# Korpora in Lehre und Forschung

Hrsg.
Andrea Abel
Renata Zanin





Freie Universität Bozen
Libera Università di Bolzano
Free University of Bozen · Bolzano

# Korpora in Lehre und Forschung

Hrsg.
Andrea Abel
Renata Zanin



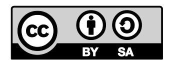




1. Auflage
www.unibz.it/universitypress

ISBN 978-88-6046-040-0
E-ISBN 978-88-6046-095-0

Design: DOC.bz



# Inhalt

# Vorwort

Das vorliegende Buch ist der Diskursgemeinschaft der am Erlernen von Fremdsprachen Interessierten zugeordnet, dazu gehören in erster Linie die Lehrpersonen, aber auch die Lernenden selbst, dazu die Experten aus Pädagogik, Didaktik, Sprachwissenschaft, Schuldirektion und Verwaltung.
Das Thema kann in seiner allgemeinsten Form als Frage nach dem Beitrag computerunterstützter Methoden beim Fremd- und Zweitsprachunterricht definiert werden. Es handelt sich um ein Thema, das seit den 80er Jahren eine Fülle von voreilig definierten Aufgaben und Lösungsversuchen hervorgebracht hat. Besonders auffallen muss die große Zahl hoch subventionierter Forschungsprojekte, die im Rahmen europäischer Fördermaßnahmen unterstützt wurden und die praktisch keinen nachweisbaren Gewinn erbracht haben. Zu diesen Momenten fehlgeleiteter, mit großem Enthusiasmus vorgetragener innovativer Vorschläge gehört auch das sogenannte Data Driven Learning (DDL), das vor 15 Jahren im Gefolge der erfolgreichen Nutzung von Korpora für lexikographische Arbeiten als Allheilmittel für den Sprachunterricht entdeckt und propagiert wurde.
Im Gegensatz zu einer solchen, von einer starken Vorentscheidung zugunsten der von technologischen Lösungsversuchen getragenen Tradition in der Methodendiskussion, ist das Projekt „Korpus Südtirol“, das von einer Forschungsgruppe EURAC-Freie Universität Bozen von 2004 bis 2007 durchgeführt wurde, als Versuch der kleinen Schritte zu charakterisieren. Ausgehend vom objektiv gegebenen Bedürfnis, die deutschen Sprachdokumente Südtirols zu sammeln und für die Forschung aufzubereiten, wurden schrittweise didaktische Anwendungen erprobt, die alle durch eine genaue Zielgerichtetheit auf spezielle Lernergruppen geprägt waren.
Ausgangspunkt für die didaktischen Anwendungen war das generell bekannte Defizit an Sprachaufmerksamkeit in Südtirol, das auf die Konstellation eines Sprachgebrauchs im Spielfeld von Dialekt, deutscher Hochsprache und italienischer Zweitsprache zurückzuführen ist. Zu dieser Beanspruchung kommt nun noch das Englische als unverzichtbare Sprache der Kommunikation.
Das Projekt, das im vorliegenden Band dokumentiert und kritisch referiert wird, trägt dieser außergewöhnlichen Situation von Sprachgebrauch, Spracherwerb und Sprachenlernen in Südtirol Rechnung.

Zentrales Anliegen des Projekts und des Bandes ist der ganzheitliche Zugang auf die Probleme des Sprachgebrauchs und des Sprachenlernens, bei dem die drei Schulsprachen Deutsch, Italienisch und Englisch, die auch die Arbeitssprachen der Freien Universität Bozen sind, als Einheit betrachtet und behandelt werden. Die Kompetenzen, die bei der Bearbeitung des Projekts eingesetzt wurden, reichen von der sprachlichen Analyse der deutschen Gegenwartssprache bis zu didaktischen Analysen von Lernprozessen beim gesteuerten Spracherwerb in der Schule und zu computertechnischen Fragestellungen zum Interface, das den Zugang zu den Daten ermöglicht.
Diese Ausrichtung und die im Rahmen des Projekts erreichten Ergebnisse geben dem Buch eine ganz eigene inhaltliche Bestimmung, die es von vergleichbaren Werken mit stärker monographischem Charakter abhebt. Es handelt sich um ein Werk, das in allen Phasen der darin beschriebenen Arbeiten und Arbeitsschritte als Ausdruck einer theoriegestützten Praxis gelten kann.
Innovative Elemente des Bandes sind: (1) die theoretische Fundierung der Lernprozesse im Rahmen eines kognitiven Modells der Grundlage der Sprachverarbeitung (das Resonanz-Modell), das gezielt verwertbar ist als Ausgangspunkt für Sprachlernprozesse mithilfe von Korpusrecherchen; (2) die innovative Nutzung von Klein- und Einzelwerkkorpora für die Sprachdidaktik als Antwort auf bisher ungenügend beantwortete kritische Fragen zur theoretischen Fundierung von korpusorientierten Sprachlernmodellen und Strategien; (3) die Anpassung bestehender Korpora an ein einheitliches, benutzerfreundliches Interface für Korpora in allen Schulsprachen, (4) die Entwicklung neuer Ansätze und computerlinguistischer Ressourcen sowie Instrumente für den (semi-)automatischen Vergleich sprachlicher Varietäten, wodurch auf Desiderate im Bereich der Sprachkontakt- und Varietätenforschung im Allgemeinen eingegangen sowie ein Beitrag zur Beschreibung der deutschen Sprache Südtirols im Besonderen geleistet wird, (5) die Präsentation von Methoden und Werkzeugen zur Ermittlung und Darstellung ausgewählter sprachlicher Phänomene, insbesondere im Bereich der syntagmatischen Relationen.
Das Buch ist ein Dokument für die Bemühungen um eine global verstandene Spracherziehung und Sprachausbildung in Südtirol, die, seit dem Konzept der integrierten Sprachdidaktik aus den 80er Jahren, als Desiderat für die Schule und die Gesellschaft anerkannt werden.

Bozen, April 2011
Andrea Abel, Renata Zanin

# Vom Interface zur Sprache und zum Spracherwerb

**Andrea Abel – Europäische Akademie Bozen**
**Renata Zanin – Freie Universität Bozen**

Jedes Wort, das verwendet wird, hinterlässt eine Spur. Wer es verwendet, zeigt, dass er/sie es gelernt hat und im Moment der Verwendung erhält das Wort die Chance, aufs Neue gelernt zu werden und so in dieser und vielleicht der nächsten Generation weiterzuleben.

Wörter überleben in Sprachen und Sprachen überleben in Sprachgemeinschaften. Sprachen sterben aus – zurzeit mit dramatischer Konstanz und Geschwindigkeit –, Wörter werden durch andere ersetzt, umgeformt, werden in neuen Gebrauchssituationen verwendet, erhalten neue Bedeutungen, werden von bestimmten Menschen oder Gruppen aufgegriffen und bevorzugt benutzt, werden aufgezeichnet, gelöscht, vergessen.

Sie überleben als gelebte Gegenwart, aber auch als Palimpsest, wobei Überschriebenes stärker wirksam wird als der ursprüngliche Text. Waren Palimpseste einst Zeichen für einen Mangel, nämlich den Mangel an wertvollem Pergament, so entstehen metaphorische Palimpsest-Wörter von heute durch Überfluss. Es sind Wörter, die vergessen werden, weil sie in der enormen Masse an Wörtern, Sätzen und Dokumenten nicht beachtet werden, weil sie zu selten sind und von der Masse der anderen Wörter überdeckt werden. Und wie die alten überschriebenen Pergamenthandschriften mithilfe von UV-Strahlen und anderen, raffinierteren Techniken wieder sichtbar gemacht werden können, so gibt es heute Algorithmen, die auch die verstecktesten Wörter aus der Masse von Trillionen von Wörtern (Tokens) aus digitalisierten Büchern herauszusuchen imstande sind. Ermöglicht wird eine solche Suche durch die N-Gram-Datenbank[1], die als Teilmenge der von Google digitalisierten Bücher im Dezember 2010 veröffentlicht und nutzbar gemacht wurde.

Mit diesem neuen Suchinstrument und der ständig wachsenden Datenbank fällt es uns leichter, uns vorzustellen, dass jedes Wort, das verwendet wird, eine Spur in der Geschichte der Menschheit hinterlässt. Und diese Vorstellung, diese Idee, war nicht nur das auslösende Moment für die Konstitution der Arbeitsgruppe, die die Google-Datenbank ins Leben gerufen hat, sondern auch dafür, eine neue Wissenschaft aus der Taufe zu heben, „culturomics“, der sogleich das Schlagwort „cultural genome“ in die Wiege gelegt wurde. Die Aufgabe und das Ziel der neuen Wissenschaft: „working to enable the quantitative

1 http://ngrams.googlelabs.com/

study of human culture across societies and across centuries“ [2]. Ihre Grundlage: Ein unerhört großes Archiv an gedruckten Zeugnissen aus der Geschichte ist digital aufbereitet und durchsuchbar gemacht worden. Die unvermeidlichen Einschränkungen können der Faszination der Datenmenge nichts anhaben. Aber die Einschränkungen sind nicht trivial: zuallererst die Einschränkung auf schriftliche Zeugnisse, dazu die weitere Einschränkung auf Gedrucktes, verbunden mit einer kontinuierlich fortschreitenden Datenerfassung – zurzeit sind 11% aller gedruckten Bücher digitalisiert. Dann kommt noch die spezielle Aufbereitung der Daten für die Suche hinzu, die natürlich nur einen Bruchteil der Gesamtdaten betreffen kann, zurzeit sind es 4%, die bevorzugt den Zeitraum von 1800–2000 betreffen und bevorzugt Dokumente, die mit einem hohen Qualitätsstandard eingescannt wurden (was Lesefehler nicht ausschließt, die schon bei den ersten Suchen auffallen). Schließlich noch eine weitere Einschränkung: Die Suche findet nicht im Archiv der Daten mit dem ursprünglichen Datenformat statt, sondern in einer eigenen N-Gram-Datenbank, die zurzeit nur Daten in den Sprachen Englisch (nach Amerikanisch, Englisch, Fiction differenziert), Chinesisch, Französisch, Spanisch, Deutsch und Russisch enthält. Auch liegt keine etwa nach Textsorten und Zeiträumen ausgewogene Datensammlung mit unterschiedlichen Filteroptionen für eine gezielte Suche vor.

Trotz aller Einschränkungen: Aktuell sind in der durchsuchbaren N-Gram-Datenbank bereits ungefähr drei Trillionen Wörter gespeichert. Zum Vergleich: Das British National Corpus BNC[3], das für lange Zeit den Standard von digital erforschbaren Datensammlungen bestimmte, verfügt über 100 Millionen Wörter. Das DWDS Kernkorpus umfasst 100 Millionen Textwörter (Tokens)[4]. Das Deutsche Referenzkorpus des Institutes für Deutsche Sprache (IDS), das COSMAS II zugänglich macht, umfasst ca. 3,2 Mrd. Wörter[5]. Die von Google aufbereiteten drei Trillionen Wörter können von allen Menschen der Welt mit Zugang zum Internet durchforscht werden. Das Programm der Forschungsgruppe an der Harvard Universität, das in Zusammenarbeit mit Google entsteht, wendet sich expressis verbis sogar an Kinder, die über ihre Sprache staunen, oder das Staunen lernen wollen, indem sie mithilfe des N-Gram-Viewers überprüfen, wie oft ein bestimmtes Wort oder eine bestimmte Wortgruppe oder Namen, Spitznamen oder welche sprachlichen Formen auch immer in der Welt vorkommen, wie sich das Vorkommen im Lauf

2 http://www.culturomics.org/cultural-observatory-at-harvard

3 http://www.natcorp.ox.ac.uk/

4 http://www.dwds.de/

5 http://www.ids-mannheim.de/cosmas2/

der Zeit verändert hat, welche Künstler öfter zitiert werden als andere, und so weiter, und so weiter.
Die Charts, die das Programm in Sekundenbruchteilen erstellt, sind in der Tat faszinierend. Geschichtlich relevante Ausdrücke können minutiös nachverfolgt werden (Suche: The Great War, World War I), aber auch zufällig Relevantes, wie das Fach „Sozialpädagogik“, das Studenten an der Freien Universität Bozen studieren, und das rasante Ansteigen der Belege dafür im letzten Jahrzehnt. Ebenso kann der Beginn der Verbreitung des Fachausdrucks „Zeigfeld“ in das Jahr 1934 datiert werden, oder „Deixis“, mit Karl Brugmann als Vorläufer um 1904. Solche rasch auffindbaren Ergebnisse, die auf einzelne Jahre eingeschränkt werden können und den direkten Zugriff auf die Quelltexte gestatten, haben nichts von einer „quick and dirty“-Informationsbeschaffung an sich. Sie verweisen mit Präzision auf Enklaven von spezialisiertem Sprachgebrauch durch Mitglieder einer wissenschaftlichen Gemeinschaft.
Der Google-Such-Algorithmus sucht nach reinen Buchstabenfolgen – N-Grams –, die Suche ist daher auf Wörter eingeschränkt; es kann nicht nach Lemmata, also den jeweiligen Grundformen, gesucht werden, und es sind keinerlei Hinweise auf semantische Differenzierungen möglich. Gerade bei sprachlich nicht geschulten Nutzern wird ein Bild von der Sprache und den Sprachen gefestigt, das auch im Schulunterricht eine feste Rolle einnimmt: Unser Bild von der Sprache ist stark von der Vorstellung vom Wort als Urelement der Sprachen geprägt. Wörter prägen sich ein, werden vergessen, stellen beim Übersetzen ein Hindernis dar und werden somit stark vom Kontext isoliert als eigenständige Elemente der Sprache wahrgenommen. Aus der Perspektive schulischen Kontakts mit Sprache kommt als zweites Grundelement die Syntax hinzu, die Syntax als Summe der Regeln, die zum korrekten Gebrauch der Wörter verhelfen. Von diesem Bild aus werden auch die Strategien zum Erwerb der Zweitsprache erfahren und bewertet, verstärkt durch den Gebrauch von Wörterbuch und Grammatik als Referenzwerke im Umgang mit Sprachen, die gelernt werden müssen.
Auf dieser Grundlage baut die neuere Forschung zur Sprache im Umfeld ihres Gebrauchs auf und entdeckt dabei, dass zusätzlich zu den Wörtern Gruppierungen von Wörtern in festen Gebrauchsformen, als eingespielte Versatzstücke, eine große Rolle spielen, und dass die Verknüpfung der Wörter zu größeren Einheiten keineswegs nur als Zusammenbauen isolierter Elemente mithilfe von Regeln vor sich geht. Ganz im Gegenteil, Sprache operiert hochgradig mit mehr oder weniger rekurrenten mehrteiligen Mustern bzw. Konstruktionen.

Diese Erkenntnisse wurden in den letzten 20 Jahren im Rahmen der Sprachforschung mit Korpusbelegen systematisch erweitert und konsolidiert, wobei nicht selten das Forschungsinteresse für Korpora und für Konstruktionen von Forschern in Personalunion vertreten wird (Stefan Gries mit einem umfangreichen Werk zu beiden Aspekten[6]).
Der zu erwartende Erfolg des N-Gram-Viewers von Google mit der Begeisterung über rasch zu erzielende plausible Ergebnisse ist ein guter Anlass, die positiven Elemente dieser in den letzten Jahrzehnten diskutierten Versuche zu neuen Theorieansätzen in ein Gesamtbild zu vereinen, mit dem Ziel, eine Diskursgemeinschaft zu bilden, in der die nach Neuerungen drängenden Eltern ebenso integriert werden wie Lehrpersonen und die für das Gedeihen der Schule verantwortlichen politischen Kräfte.
Ein Ziel, das für das Wachsen dieser Diskursgemeinschaft angestrebt werden muss, ist, ein gemeinsames Bild der Sprachen in ihrer Funktionalität zu gewinnen, ein Bild, das wissenschaftlich abgesichert ist, aber darüber hinaus auch anschaulich die Wege in die Praxis des Sprachkontakts und des Spracherwerbs aufzeigen kann.
Es beginnt damit, die Grundanschauung, die wir von der Sprache haben, kritisch zu überprüfen. Niemand, der Sprache und Grammatik in der Schule gelernt hat, wird leugnen, dass „Wörter Einzelelemente sind, die mit Hilfe des Wissens um die Grammatik zusammen gesetzt werden". Wir könnten mit Lessing antworten: „Ich bin dieser Niemand", und zwar in Übereinstimmung mit einer großen Zahl von Vertretern unterschiedlicher Richtungen der Sprachwissenschaft, die zu einer Sicht der Sprache gefunden haben, in der Wort und Syntax nicht als getrennt angesehen und behandelt werden. Diese neue Sicht der Sprache(n) hat weitreichende Folgen für die Sprachwissenschaft, inklusive der angewandten Sprachwissenschaft in den Bereichen des Spracherwerbs, des Übersetzens und der Sprachpflege im Umfeld institutionellen Sprachgebrauchs.
In all diesen Bereichen ist in einem bestimmten Stadium des Erlernens von „skills" die Aufmerksamkeit zu schulen, die Wahrnehmung von Mustern zu entwickeln und zu üben. Dazu ist, wie unzählige Studien, Theorieansätze und die Erfahrung vieler Sprachlerner in der Kontaktsituation einer mehrsprachigen Lebensumwelt zeigen, weit mehr als ein unbeteiligtes Hinschauen auf die sprachlichen Elemente nötig, die vor unsere Augen (oder Ohren) treten. Sprache muss voll und ganz verstanden werden und in einem als sinnvoll erfahrenen Kontext der eigenen Lebenswelt als wirksam erkannt werden.

6 http://www.linguistics.ucsb.edu/faculty/stgries/

Nicht die unentwegte Sprachberieselung ist entscheidend, sondern das Aufnehmen voll verstandener Sprache in sinnvollen Lebenskontexten. Für den Zweitsprachunterricht hat das Lily Wong Fillmore exemplarisch in ihrer oft zitierten Dissertation ausgesprochen: "Exposure to language is not enough; it must be directed at the learner, and shaped with his needs and abilities in mind." (Fillmore 1976[7], 119). In der neueren Forschung spricht man von „intake" im Gegensatz zum rein mechanischen „input", der auch ohne Spuren zu hinterlassen wieder verschwinden kann. Diese Beobachtungen, Erfahrungen und Theorieansätze zum Fremdsprachenlernen und zum Zweitspracherwerb stellen einen möglichen Hintergrund für den Einsatz von Korpusinstrumenten im Sprachunterricht dar.
Dazu kommt in einer mehrsprachig orientierten Gemeinschaft wie der in Südtirol – in diesem Kontext wählen wir den italienischen Ausdruck „comunità di pratica", der metonymisch dem deutschen Ausdruck „Diskursgemeinschaft" entspricht – der Vorteil, dass unsere Lernenden sinnvolle Suchen/Korpusrecherchen in den drei Schulsprachen durchführen und sehr rasch auch tatsächlich selbständig durchführen können. Sehr oft ist es gerade die unerwartete Herausforderung, das gerade erst entdeckte Erkenntnispotential des neuen Instrumentes für individuelle Fragestellungen in einer Sprache auch für andere Sprachen zu nutzen und selbständig analoge Suchanfragen in mehreren Sprachen zu konfigurieren, die zum Auslöser für ein nachhaltiges Interesse an Sprache, an Sprachformen und -mustern wird.

7 Fillmore, Lily Wong (1976): The second time around: cognitive and social strategies in second language acquisition, Ph.D. dissertation, Stanford.

# Korpus Südtirol – Aufbau und Abfrage

**Stefanie Anstein – Europäische Akademie Bozen**
**Margit Oberhammer – Freie Universität Bozen**
**Stefanos Petrakis – Universität Zürich**

## Abstract

Dieser Beitrag gibt einen Einblick in die Initiative *Korpus Südtirol* und die ca. 70 Millionen annotierte Tokens umfassende Textsammlung sowie in deren Anwendungsmöglichkeiten. Der inhaltlichen Beschreibung der Textbasis folgt eine Erklärung der technischen Arbeitsschritte zur Korpusaufbereitung und die Vorstellung der Benutzeroberfläche. Das *Korpus Südtirol* dokumentiert die geschriebene Sprache des 20. Jahrhunderts in Südtirol und dient als Instrument für linguistische Untersuchungen.

## 1. Einführung

Die deutsche Sprache ist in der Autonomen Provinz Bozen / Südtirol (Italien) der italienischen gleichberechtigte Amtssprache, festgelegt im Autonomiestatut[1] (Art. 99, 100; erstes Abkommen: Pariser Vertrag 1946). In den ladinischen Tälern kommt Ladinisch als dritte Amtssprache hinzu. Von den ca. 500.000 Einwohnern Südtirols sind etwa zwei Drittel deutscher Muttersprache, ein Viertel italienischer und der Rest ladinischer Muttersprache. Deutsch wird in nahezu allen Bereichen des öffentlichen Lebens parallel zum Italienischen verwendet - in den Medien, in Schulen, Weiterbildungsinstituten oder in der öffentlichen Verwaltung.

Die deutsche Sprache weist Unterschiede zwischen ihren Varietäten in Deutschland, Österreich, der Schweiz sowie in weiteren sogenannten nationalen Halbzentren wie Südtirol auf (vgl. Ammon 1995a und 1995b). Diese Unterschiede bestehen auf verschiedenen Ebenen, von der Aussprache- über die Wort- bis zur Satzebene. Wortbeispiele für die Südtiroler Varietät sind *Kondominium*, *Halbmittag*, *Proporz*, *Waal* oder *Maturant* (vgl. Ammon et al. 2004 und Abfalterer 2007).

Während eine historische und gesellschaftspolitische Auseinandersetzung mit der Vergangenheit in Südtirol in den 80er Jahren des vergangenen Jahrhunderts eingesetzt hat, gibt es im sprachlichen Bereich noch große Lücken. Was Wolfgang Klein für den gesamten deutschsprachigen Raum feststellt, die weitgehend im Dunkeln liegenden Sprachverhältnisse in der ersten Hälfte des 20. Jahrhun-

---

1 <http://www.provinz.bz.it/land/landesverwaltung/suedtirol/autonomiestatut.asp> (Stand: Jänner 2011)

derts, trifft für Südtirol in besonderem Maße zu. Auch hier ist der unzureichende Forschungsstand nicht zuletzt „ein Indiz für die mangelnde Fähigkeit, sich den Texten der Vergangenheit zu stellen und die Sprache dieser Texte als das Fundament der Gegenwartssprache zu begreifen“[2].
Die Textsammlung des *Korpus Südtirol*[3] soll zur Auseinandersetzung mit der sprachlichen Vergangenheit anregen, mit der Geschichte der deutschsprachigen Schriftlichkeit des 20. Jahrhunderts in Südtirol. Ziel ist u.a. ein besseres Verständnis des gegenwärtigen Sprachgebrauchs.

Das *Korpus Südtirol* dokumentiert die geschriebene deutsche Sprache in Südtirol und hat eine wichtige Archivfunktion. Es wird für linguistische Untersuchungen (s. auch Beitrag Abel/Anstein in diesem Band) sowie für Sprachaufmerksamkeitsförderung und Didaktik (s. Beitrag Zanin in diesem Band) verwendet. Die mit Metainformationen angereicherten Texte stellen zudem eine Basis für philologische und stilistische Vergleiche dar, für Beobachtungen von Sprachwandelphänomenen und von textsortenspezifischem Sprachgebrauch.
Das Korpus Südtirol beteiligt sich an der Initiative *C4*[4], die vier der deutschen Varietäten in Europa zusammenfassend dokumentiert und den öffentlichen Zugang zu vergleichbaren Korpora der folgenden Arbeitsstellen anbietet: *DWDS*[5] *- Das Digitale Wörterbuch der deutschen Sprache des 20. Jh.* der Berlin-Brandenburgischen Akademie der Wissenschaften, *AAC*[6] *- Austrian Academy Corpus* der Österreichischen Akademie der Wissenschaften in Wien, *Schweizer Text Korpus*[7] des Deutschen Seminars der Universität Basel.

## 2. Textbasis und Metadaten

Um eine ausgeglichene und möglichst repräsentative Textsammlung zu erhalten, werden in Anlehnung an das DWDS-Korpus vier Textsorten zu gleichen Teilen für das 20. Jahrhundert bis heute gesammelt: Belletristik, Gebrauchstexte, journalistische Prosa und Sachtexte.[8]

2 Klein, Wolfgang: <http://www.bbaw.de/forschung/dwds/projektdarstellung > (Stand: Jänner 2011)

3 s. auch <http://www.korpus-suedtirol.it> (Stand: Jänner 2011) für weitere Informationen, Literaturangaben sowie den Link zur Suche in verschiedenen Teilkorpora

4 <http://www.korpus-c4.org> (Stand: Jänner 2011)

5 <http://www.dwds.de> (Stand: Jänner 2011)

6 <http://www.aac.ac.at> (Stand: Jänner 2011)

7 <http://www.dwds.ch> (Stand: Jänner 2011)

8 zur zeitlichen Planung s. Kapitel 3.

Die Quantität, Qualität und zeitliche Verteilung muss vor dem Hintergrund der geschichtlichen Situation Südtirols im 20. Jahrhunderts gesehen werden.[9] Bis zum Ende des Ersten Weltkriegs ist Südtirol mit 250.000 deutschsprachigen BewohnerInnen Teil Tirols und gehört somit zu Österreich, beziehungsweise zur späten Donaumonarchie. Literarisches und wissenschaftliches Zentrum am Beginn des 20. Jahrhunderts und häufiger Publikationsort der Druckerzeugnisse ist Innsbruck. Nach der Abtrennung von Österreich und während des erstarkenden italienischen Faschismus gehen in den 20er Jahren die schriftlichen Quellen bis auf einige wenige Erzeugnisse in Kleinstverlagen völlig zurück. Gab es 1925 noch drei Tageszeitungen und eine Vielzahl verschiedener Wochen- und Monatszeitschriften, so wurden diese im Jahr darauf durch das faschistische Pressegesetz verboten. In den späten 30er und in der ersten Hälfte der 40er Jahre fördert die Politik des Nationalsozialismus vor allem eine bestimmte Sorte des regionalen Schrifttums, die Heimatliteratur. Die Romane von Joseph Georg Oberkofler, Franz Tumler, Maria Veronika Rubatscher und Luis Trenker erscheinen in hohen Auflagen in Verlagen in Deutschland. In der Nachkriegszeit hinterlassen die beiden Diktaturen eine Bildungskatastrophe, die sich bis weit in die 60er Jahre hinein bemerkbar macht. Zudem sind nach dem Zweiten Weltkrieg Politik, Medien und Literatur eng miteinander verflochten und das Verlags- und Pressewesen ist monopolisiert. Publiziert werden wiederaufgelegte Texte aus den 30er Jahren, Kalender- und religiöse Literatur. Kulturelle Institutionen bemühen sich vor allem um die Pflege der volkstümlichen Kultur und deren sprachlichen Erzeugnissen. Seit der Durchsetzung der Autonomie in den 70er Jahren und dem wirtschaftlichen Aufschwung seit den 80er Jahren blühen Presse- und Verlagswesen. Die deutsche Sprache gewinnt verlorengegangene Verwendungsdomänen zurück und ihr Prestige wird gehoben. In den schriftlichen Quellen zeigen sich Problembewusstsein und Interesse an der kritischen Auseinandersetzung mit Vergangenheit und Gegenwart. Das inhaltliche und sprachliche Spektrum der Texte wird breiter.

## 2.1 Kriterien für die Zusammensetzung

Die Zusammensetzung der Textbasis lehnt sich an jene des Kernkorpus des *Digitalen Wörterbuchs der deutschen Sprache* und an das *Schweizer Text Korpus* an.

### 2.1.1 Formale Kriterien

Aufgenommen werden Druckwerke und maschinenschriftliche Texte (publizierte Diplomarbeiten und Dissertationen) in Prosa. Ein wichtiges Kriterium

9 s. auch bei Steininger, Rolf: <http://www.rolfsteininger.at> (Stand: Jänner 2011)

für die Aufnahme ist der Status der Urheberrechte. Die Textsammlung beschränkt sich in erster Linie auf Standardsprachliches. Dazu zählen auch Texte in stilisiertem standardnahem Dialekt, wie er in Dramen einiger Südtiroler AutorInnen verwendet wird. Das *Korpus Südtirol* ist formal wie erwähnt in die vier Kategorien Belletristik, Sachtexte, Journalistische Prosa und Gebrauchstexte eingeteilt. Es geht bei der Einteilung weniger um eine genaue Texttypologie als vielmehr um eine große Bandbreite an Textsorten. Eine ausgewogene Verteilung dieser Textsorten wird nach Möglichkeit angestrebt. Bei Texten, deren Verfasser bekannt sind, sind dies AutorInnen, die in Südtirol geboren und/oder den Großteil ihres Lebens dort verbringen oder als RedakteurInnen, HerausgeberInnen und PublizistInnen einen maßgeblichen Einfluss auf die geschriebene Sprache haben.

| | |
|---|---|
| Belletristik | Beinhaltet möglichst viele unterschiedliche Gattungen, ausgenommen Lyrik (Romane, Novellen und Erzählungen, Biographien, literarische Briefe, Essays, Sagen, Märchen, Städtebilder, Heimatliteratur, Kriminalgeschichten, Trivialliteratur, Kinderliteratur, Kalenderliteratur, Anthologien, Dramentexte).<br>Aufgenommen in diese Kategorie wurde auch autobiographische Literatur, die in Südtirol eine wichtige Rolle spielt und für sprachliche Untersuchungen vor allem im Bereich der Varietätenlinguistik aufschlussreich ist: Tagebücher, Kindheitserinnerungen, Kriegserinnerungen, Reisebeschreibungen, Bergsteiger- und Expeditionsberichte. |
| Journalistische Prosa | Südtirol weist im 20. Jahrhundert eine sehr wechselvolle Zeitungsgeschichte auf mit vielen Neugründungen, Verboten, Unterbrechungen Wiederaufnahmen, kurzlebigen Blättern. Die Aufnahme der Tages- und Wochenzeitungen geschieht in enger Kooperation mit der *Landesbibliothek Dr. Friedrich Teßmann,* die bereits die meisten historischen Südtiroler Zeitungen digitalisiert hat (z.B. *Brixner Chronik, Der Tiroler, Bozner Zeitung, Bozner Nachrichten, Burggräfler, Pustertaler Bote, Der Pfeil, Tandem, Dolomiten…*). Zusätzlich Magazine (*ff, IN Südtirol, Südtiroler Wirtschaftszeitung*), Studentenzeitungen (*Der fahrende Skolast, Quer*), unregelmäßig erscheinende Zeitungen von Vereinen und Verbänden |

(Theaterverband, Sängerbund, Gastwirteverband, Imkerverband, Handwerkerverband), Veröffentlichungen des Landespresseamtes, Bezirks- und Gemeindeblätter.

Sachprosa
Texte aus verschiedenen Fachgebieten: populärwissenschaftliche Aufsätze in Zeitschriften (Der Schlern, Arunda), in Jahrbüchern und Tagungsbänden ebenso wie umfangreichere, im engeren Sinn wissenschaftliche Texte (auch Diplomarbeiten und Dissertationen). Außerdem Monographien und themenspezifische Zeitschriften (Geschichte und Region, Theologisches Forum, Zeitschrift des Burgeninstituts- arx, forum- Schule heute).

Gebrauchstexte
Als Gebrauchstexte wurden „Texte für jemanden“ kategorisiert (Schulbücher, Gesundheitsratgeber, Kochbücher, Wanderführer, Gebietsführer, Gebrauchsanweisungen, Werbetexte, Programmhefte, Verordnungen).

#### 2.1.2. Zeitliche Kriterien

Die Textgrundlage soll möglichst repräsentativ das geschriebene Südtiroler Deutsch des 20. Jahrhunderts erfassen und wird zudem regelmäßig durch Texte aus dem 21. Jahrhundert ergänzt und erweitert. Der Textbestand wird wie im Korpus des *DWDS* in Dekaden gegliedert. Ein zentrales Kriterium für die Aufnahme eines Textes ist das Publikationsjahr.

#### 2.1.3. Inhaltliche Kriterien

Es werden Texte aus sehr unterschiedlichen Themengebieten gesammelt, aus den Bereichen: Philologie, Pädagogik, Rechtskunde, Wirtschaftswissenschaften, Technik, Musik- und Theaterwissenschaften, Medizin, Philosophie, Landeskunde, Volkskunde, Geschichte, Geografie, Geologie, Botanik, Kunstgeschichte, Archäologie, Theologie. Die inhaltliche Erschließung erfolgt mit dem Dewey-Dezimalklassifikationssystem[10], der international am weitesten verbreiteten Klassifikation für Bibliotheksbestände.

### 2.2. Quellen für die Metadaten

Die bibliographische Situation in Südtirol ist insgesamt eine besondere, weil es ähnlich wie für Nordtirol für das 20. Jahrhundert keine fortlaufende bibliographische Berichterstattung gibt und die bibliographischen Daten aus unterschiedlichen analogen und digitalen Quellen recherchiert werden müssen. Im

10 <http://www.ddc-deutsch.de> (Stand: Jänner 2011)

Folgenden werden die wichtigsten Lexika, Datenbanken, Bibliotheken und Internetarchive[11] aufgelistet.

### 2.2.1. Belletristik

- <http://www.literature.at> (*alo – austrian literature online*)
- <http://www.biographien.ac.at> (Österreichisches Biographisches Lexikon - Online Edition)
- <http://www.uibk.ac.at/brenner-archiv> (Literaturdatenbank)
- <http://www.kuenstlerbund.org/de> (Dokumentationsstelle für Neuere Südtiroler Literatur)
- ausgewählte deutschsprachige Buchveröffentlichungen Südtiroler AutorInnen seit 1990: In: (W)orte. Zeitgenössische Literatur aus und über Südtirol (2005); herausgegeben, kommentiert und übersetzt von Siegrun Wildner. Innsbruck-Bozen-Wien: Skarabaeus
- Gruber, Alfred (Hrsg.)(1990): Nachrichten aus Südtirol. Deutschsprachige Literatur in Italien. (=Auslandsdeutsche Literatur der Gegenwart. Band 4.), Hildesheim: Verlag Georg Olms AG
- Enzinger, Moritz: Die deutsche Tiroler Literatur bis 1910 - Ein Abriß. Wien, 1929
- Deutsches Literatur-Lexikon (1968 ff), 3. Auflage. Bern, München, Stuttgart: Francke.
- Oberschelp, Reinhard (Hrsg.)(1976): Gesamtverzeichnis des deutschsprachigen Schrifttums (GV). 1911–1965. München, u.a.: Saur.

### 2.2.2. Journalistische Texte

- <http://www.provinz.bz.it/lpa/medienverzeichnis.asp> (Online-Medien, Bezirksblätter, Gemeindeblätter, Monatsschriften)
- <http://www.press-guide.com/italy.htm> (deutschsprachige Tages- und Wochenzeitungen)
- <http://www.provinz.bz.it/lpa> (Landespresseamt)
- Demattia, Roman: Verzeichnis der in Südtirol erschienenen Zeitschriften und Periodika von 1945 bis 1978 (einschließlich solche der Provinz Trient, soweit sie sich auf die gesamte Region Trentino-Südtirol beziehen), die in Bibliotheken Südtirols einschließlich der Universitätsbibliothek Innsbruck und des Tiroler Landesmuseums *Ferdinandeum* vorhanden sind. (Wissenschaftliche Hausarbeit) Bozen 1980.

11 Stand jeweils: Jänner 2011

- Kugler, Dorothea: Die Südtiroler Periodika bis 1945: Standortnachweis und statistische Auswertungen. (wissenschaftliche Hausarbeit) Bozen 1988.

#### 2.2.3. Sach-/Fach- und Gebrauchsprosa

- < http://www.tessmann.it> (Tirolensien und digitales Zeitungsarchiv)
- <http://www.uibk.ac.at/ulb/information/bibliographie.html#tirolensien> (Tirolensienabteilung der Universitätsbibliothek Innsbruck)
- Buchland Tirol 1980–1990 (1991) Südtirol, Nordtirol, Osttirol. Literatur-Handbuch und Tirolensienverzeichnis; hrsg. v. Südtiroler Landesregierung und Amt der Tiroler Landesregierung. Bearbeiter: Ferruccio Delle Cave, Kurt Habitzel, Anton Sitzmann, Toni Unterkircher. Wissenschaftliche Beratung: Johann Holzner. Bozen-Innsbruck: Amt der Tiroler Landesregierung und Südtiroler Landesregierung.
- Der Schlern. Monatszeitschrift für Südtiroler Landeskunde. Verfasser-, Personen-, Orts-, Sach- und Bildregister: Register für die Jahrgänge 1920–1978 (1979). Bearbeiter: Hans Griessmair. Bozen: Athesia.
- Pfaundler-Spat, Gertrud (2005): Tirol Lexikon. Ein Nachschlagewerk über Menschen und Orte des Bundeslandes Tirol. Innsbruck Wien Bozen: Studienverlag
- Habitzel, Kurt: Literatur und Literaturbetrieb in Tirol im Zeichen des Anschlusses. Innsbruck 1990 (Diplomarbeit)
- Auer, Werner, Kurt Gamper (1984): Tirol, schöpferisches Land. Steiger Verl., Innsbruck: Steiger Verlag
- Archivkatalog des Universitätsverlags A. Weger Brixen
- Archivkatalog des Verlags Athesia Bozen
- Bibliothekskatalog des Tiroler Landesmuseums Ferdinandeum Innsbruck

## 3. Datenbank und Aufbereitung

In der Bestandsaufnahme-Datenbank des Korpus Südtirol befinden sich ca. 294.000 Einträge mit Metadaten, die manuell und automatisch erstellt wurden. Jeder Eintrag beschreibt ein „Werk“, d.h. einen Text mit einem eindeutigen Titel und einem bestimmten Autor. Abbildung 2 zeigt eine Übersichtsstatistik der in der Bestandsaufnahme befindlichen Werke an Hand von Seitenzahlen.

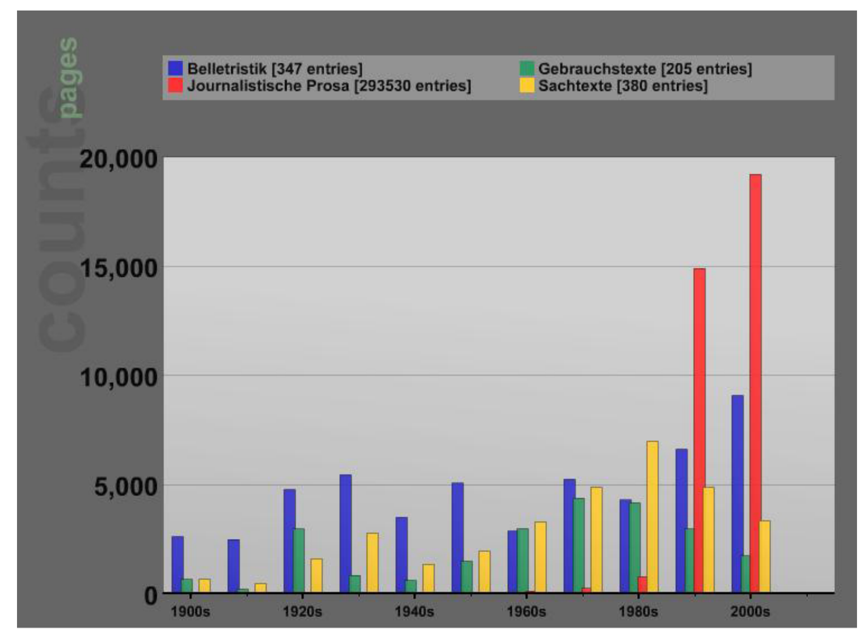



Abb. 2: Übersicht der Einträge in der Bestandsaufnahme des Korpus Südtirol, sortiert nach Textsorten und Dekaden, dargestellt auf der Grundlage von Seitenzahlen

Zur Erstellung einer formatierten Bibliographie wurde ein Werkzeug entwickelt, mit dem die gewünschten Texttypen, Dekaden und die Art der Sortierung ausgewählt werden kann.

Zu jedem Eintrag in der Bestandsaufnahmendatenbank werden halbautomatisch eine Reihe von Metadaten erfasst. Ein Auszug ist in Abb. 4 zu sehen.

| | |
|---|---|
| SourceTitle | Tschonnie Tschenett - Grobes Foul |
| SourceSubtitle | Kriminalroman |
| ConversionType | analog |
| SourceExtentPagecount | 189 |
| SourceLocation | Tessmann-Bibliothek |
| SourcePublicationISBN | 3-85218-148-8 |
| SourcePublicationWhere | Innsbruck |
| SourcePublicationWho | Haymon |
| SourcePublicationYear | 1993 |
| TaxonomyPublicationType | Buch |
| TaxonomySourceFormat | analog |
| TaxonomyOriginalTypeface | Antiqua |
| TaxonomyTextCategory | Belletristik |

Abb. 4: Auszug aus der Datenbank zur Erfassung der detaillierten Metadaten für die Bestandsaufnahmeneinträge

Jeder Datenbankeintrag, der sich wie erwähnt auf ein Werk (Artikel, Buch etc.) bezieht, wird mit dem Eintrag zu seinem Autor verlinkt. Abb. 5 zeigt das Beispiel zum vorher beschriebenen Bucheintrag.

| | |
|---|---|
| FullName | Kurt Lanthaler |
| Surname | Lanthaler |
| ForeName | Kurt |
| YearOfBirth | 1960 |
| YearOfDeath | - |
| BirthPlace | Bozen |

Abb. 5: zum Eintrag in Abb. 4 gehörige Autorenmetadaten

Die Informationen in der Datenbank werden mit Hilfe eines speziellen Programms in ein nach den Richtlinien der *Text Encoding Initiative*[12] gültiges XML[13]-Dokument umgewandelt. Ein Ausschnitt aus dem vorliegenden Beispiel ist in Abb. 6 zu sehen.

```
– <teiHeader id="id1191458784723">
   – <fileDesc>
      – <titleStmt>
         – <title>
              Tschonnie Tschenett - Grobes Foul [Machine readable version]
           </title>
         – <author>
            – <persName>
                 <name type="full">Kurt Lanthaler</name>
                 <foreName>Kurt</foreName>
                 <surname>Lanthaler</surname>
              </persName>
           </author>
         – <respStmt>
```

Abb. 6: zum Eintrag in Abb. 4 gehöriges XML-Dokument

Teilweise liegen Texte bereits in elektronischer Form vor; ein Großteil muss jedoch zunächst digitalisiert werden. Nach der erfolgten Digitalisierung werden die XML-Daten von Text und Metadaten-Header zusammengefügt und ergeben somit eine mit Metadaten angereicherte Textbasis, aus der an Hand verschiedenster Kriterien Subkorpora halbautomatisch erstellt werden können.

Bisher wurden rund 70 Millionen laufende Textwörter vollständig aufbereitet und abfragbar gemacht. Ein Großteil davon (ca. 66 Millionen Wörter) besteht aus Texten der Tageszeitung *Dolomiten* der Jahre 1991–2006, die uns der Verlag *Athesia* mit einem Nutzungsvertrag zur Verfügung gestellt hat. Der Rest beinhaltet verschiedene belletristische Werke oder auch neuere Zeitschriften. In einem knapp 2 Millionen Textwörter beinhaltenden Subkorpus steht ein zeitlich und textsortenbezogen ausgeglichenes Korpus zur Abfrage zur Verfügung, dessen Anreicherung über die nächsten Jahre weiterlaufen soll.

12 <http://www.tei-c.org/release/doc/tei-p4-doc/html> (Stand: Jänner 2011)

13 Extensible Markup Language; <http://www.w3.org/TR/xml11> (Stand: Jänner 2011)

Die Texte wurden tokenisiert, d.h. in kleinste sinnvolle Einheiten auf Wortebene (Wörter, Satzzeichen) zerlegt, und linguistisch mit Wortart- und Lemma-Informationen annotiert, d,h, angereichert (TreeTagger, Schmid 1994). Der Chunker YAC (Kermes 2003) diente teilweise zur weiteren Annotation von Chunks (grammatikalisch zusammengehörige Elemente eines Satzes; s. Abney 1991) und morphosyntaktischen Annotationen, z.B. Kasus und Numerus von Substantiven.[14]
Verwendete Abfragesysteme zum Auffinden von bestimmten Lemmata oder Mustern in annotierten und aufbereiteten Korpora sind die Korpuswerkzeuge des Instituts für Maschinelle Sprachverarbeitung (IMS): CorpusWorkBench (Christ 1994) und Corpus Query Processor CQP (Evert 2005), auf denen die eigens erstellte Benutzeroberfläche aufbaut. Für interne Anwendungen kommt auch die Sketch Engine (Kilgarriff et al. 2004) zum Einsatz.

## 4. Abfrageoberfläche und Benutzerzugang

Die Registrierung zur Benutzung der Abfrageoberfläche kann unter <http://www.korpus-suedtirol.it>, Menüpunkt *Suche in den Korpora,* durchgeführt werden.
In dieser Benutzeroberfläche können drei verschiedene Arten von Suchoberflächen verwendet werden. Die *Einfache Suche* ist für Einsteiger ohne Korpuserfahrung konzipiert, die *Erweiterte Suche* spricht Nutzer an, die bereits Erfahrung mit Korpusarbeit und linguistisches Wissen haben und über die *Expertensuche* können Anfragen in der CQP-Spezialsyntax gestellt werden.
Einige Beispiele für verschiedene Anfragen werden im Folgenden dargestellt.
Für Einzelwortsuchen ist die *Einfache Suche* ausreichend. Ihr Suchfenster ist ähnlich einer Internetsuchmaschine zu benutzen. Ein Beispiel ist in Abb. 7 und 8 zu sehen.

14 Für weitere Details zur Korpusaufbereitung s. Lemnitzer/Zinsmeister 2006.

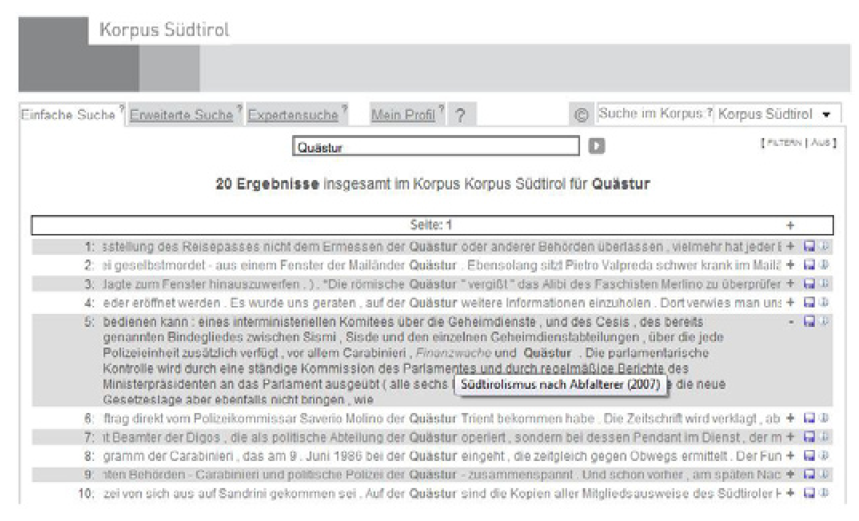



Abb. 7: Anfrage an das Korpus Südtirol mit der Option *Einfache Suche*

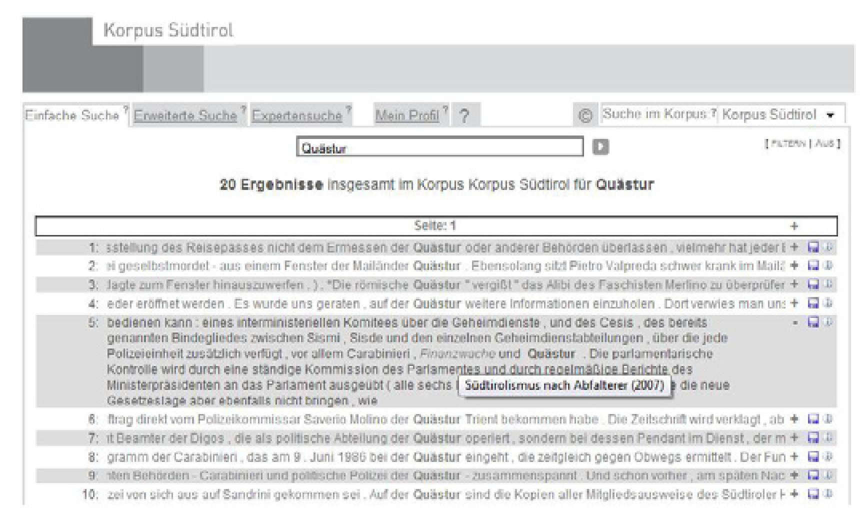



Abb. 8: Ergebnis der Anfrage aus Abb. 7

Mit der *Erweiterten Suche* können komplexere Anfragen gestellt werden, die auch die Annotationen in die Suche mit einbeziehen, z.B. „alle Adjektive, die vor dem Wort *Stimmzettel* vorkommen" (Abb. 9).




Abb. 9: Anfrage an das Korpus Südtirol mit der Option Erweiterte Suche

Die *Expertensuche* erlaubt Anfragen mit regulären Ausdrücken nach der CQP-Syntax, wenn z.B. alle Nomina mit dem Präfix „Vor-“ gefunden werden wollen (Abb. 10).

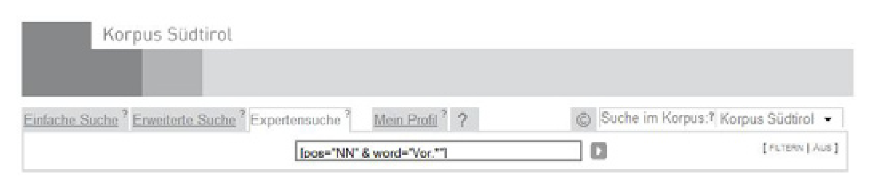



Abb. 10: Anfrage an das Korpus Südtirol mit der Option *Expertensuche*

## 5. Ausblick

Im weiteren Fortgang der Initiative ist eine umfassende Korpusanreicherung auf Grund der erstellten Bestandsaufnahme geplant, um die Sprachdokumentation, -untersuchung und -didaktik in Südtirol zu unterstützen. Zusätzlich sollen weitere Spezialkorpora hinzugefügt werden, z.B. Lernertexte und parallele Korpora. Die Annotationen werden erweitert, auch durch die Anpassung der Annotationswerkzeuge an die Südtiroler Varietät. Des Weiteren werden Werkzeuge zum systematisierten Vergleich von Varietätenkorpora weiterentwickelt (Abel/Anstein in diesem Band und Anstein 2009) und umfassende Analysen durchgeführt.

## Literaturverzeichnis

Abfalterer, Heidemaria (2007): *Der Südtiroler Sonderwortschatz aus plurizentrischer Sicht. Lexikalisch-semantische Besonderheiten im Standarddeutsch Südtirols.* Innsbruck: University Press.

Abney, Steven (1991): *Parsing By Chunks*. In: Berwick, Robert / Abney, Steven / Tenny, Carol: Principle-Based Parsing. Dordrecht: Kluwer Academic Publishers.

Ammon, Ulrich (1995a): *Vorschläge zur Typologie nationaler Zentren und nationaler Varianten bei plurinationalen Sprachen – am Beispiel des Deutschen.* In: Muhr, Rudolf / Schrodt, Richard / Wiesinger, Peter (Hgg.): Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachliche Aspekte einer nationalen Variante des Deutschen. Wien: Hölder-Pichler-Tempsky, 111-120.

Ammon, Ulrich (1995b): Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. Berlin-New York: de Gruyter.

Ammon, Ulrich / Bickel, Hans / Ebner, Jakob / Esterhammer, Ruth / Gasser, Markus / Hofer, Lorenz / Kellermeier-Rehbein, Birte / Löffler, Heinrich / Mangott, Doris / Moser, Hans / Schläpfer, Robert / Schloßmacher, Michael / Schmidlin, Regula / Vallaster, Günter (2004): *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol.* Berlin-New York: de Gruyter.

Anstein, Stefanie (2009): *Vis-À-Vis - a System for the Comparison of Linguistic Varieties on the Basis of Corpora.* In: Lyding, Verena (Hg.): Proceedings of LULCL II 2008. Bozen: EURAC 59-64.

Christ, Oliver (1994): *A Modular and Flexible Architecture for an Integrated Corpus Query System.* In: Proceedings of COMPLEX 1994, Ungarn, 23-32.

Evert, Stefan (2005): *The CQP query language tutorial.* Technical report, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
<http://cwb.sourceforge.net> (Stand: Jänner 2011)

Kermes, Hannah (2003). *Off-line (and On-line) Text Analysis for Computational Lexicography*. Dissertation: Universität Stuttgart.

Kilgarriff, Adam / Rychlý, Pavel / Smrz, Pavel / Tugwell, David (2004): *The Sketch Engine.* In: Williams, Geoffrey / Vessier, Sandra (Hgg.): Proceedings of the Eleventh Euralex International Congress. Lorient: UBS, 105-115.

Lemnitzer, Lothar / Zinsmeister, Heike (2006): *Korpuslinguistik*. Tübingen: Gunter Narr Verlag.

Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees.* In Proceedings of the International Conference on New Methods in Language Processing. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger> (Stand: Jänner 2011)

# Korpus Südtirol
# Varietätenlinguistische Untersuchungen

**Andrea Abel – Europäische Akademie Bozen**
**Stefanie Anstein – Europäische Akademie Bozen**

Abstract
Ein Überblick über relevante Arbeiten zur deutschen Sprache in Südtirol seit den 60er Jahren und davon abgeleitete Forschungsdesiderate bilden im folgenden Beitrag den Ausgangspunkt für eine systematische Untersuchung und Erfassung von möglichen Besonderheiten des Deutschen in Südtirol auf verschiedenen sprachlichen Ebenen, für die nun erstmals eine durch computerlinguistische Werkzeuge aufbereitete Datensammlung zur Verfügung steht: das im Entstehen begriffene Korpus Südtirol[1]. Vor diesem Hintergrund werden methodische Ansätze sowie erste linguistische Untersuchungen und eine Reihe von Fallstudien vorgestellt und erläutert. Sie bieten wertvolle Anhaltspunkte für vertiefende Untersuchungen und die (Weiter-)Entwicklung von Werkzeugen zum semiautomatischen Varietätenvergleich. Auf diese Weise ist es möglich, bisherige Ergebnisse aus der Sprachkontakt- und Varietätenforschung zu untermauern sowie zu ergänzen und – unter Rückgriff auf aktuelle Verfahren und Methoden – einen Beitrag zur Beschreibung der deutschen Sprache Südtirols zu leisten.

## 1. Einleitung

Die deutsche Sprache in Südtirol ist in vielerlei Hinsicht interessant für vergleichende linguistische Untersuchungen. Forschungsbedarf ergibt sich dabei zum einen aus der marginalen Lage des nationalen Halbzentrums Südtirol (zum Begriff siehe Ammon 1995a und 1995b) im Hinblick auf die Varietäten des plurizentrischen Deutschen, zum anderen aus der Sprachkontaktsituation bzw. den vielschichtigen Diglossie- und Polyglossiesituationen, die sowohl den Gebrauch als auch die Entwicklung der Sprache(n) im mehrsprachigen Kontext maßgeblich bestimmen. Veränderungen im Sprachgebrauch und Sprachwandel gehören zur inneren Dynamik von Sprachen. Sprachliche Normen sind demgegenüber stabiler, aber auch sie unterliegen Veränderungen in der Zeit, und damit verbunden sind Veränderungen in der Beurteilung und Be-

1 <http://www.korpus-suedtirol.it> (Stand: Jänner 2011)

wertung dessen, was ihnen entspricht. Was zunächst noch verbindlich ist, stellt sich später als Abweichung von Normen und Konventionen dar (vgl. Sitta 2001; Sieber/Sitta 1994). Daher gilt es, regelmäßig und systematisch zu beobachten und zu untersuchen, ob in einer Sprachgemeinschaft sprachliche Verluste und Defizite festzustellen sind und worauf sie sich zurückführen lassen.
Verschiedene Aspekte der deutschen Standardsprache in Südtirol sind in den vergangenen Jahrzehnten in einer Vielzahl von Einzelstudien und Publikationen untersucht und beschrieben worden. Doch hat sich nach einer Analyse der bisherigen Arbeiten und methodischen Ansätze, die im kommenden Abschnitt überblicksartig vorgestellt werden, eine Forschungslücke klar herauskristallisiert (vgl. dazu auch Abel/Anstein/Ties 2008: 243 f.): Ein Desiderat bilden demnach groß angelegte, systematische Untersuchungen zu den Besonderheiten der deutschen Schriftsprache in Südtirol mithilfe neuer korpuslinguistischer Methoden, und zwar einerseits auf lexikalischer Ebene (Einzel- und Mehrwortebene) zur Untermauerung und Ergänzung bisheriger Forschungsergebnisse. Andererseits sind systematische Studien auf syntagmatischer (u.a. Analyse rekurrenter Sprachmuster, Kollokationen) bis hin zur textuellen Ebene (u.a. Analyse der Textstruktur) notwendig, um Besonderheiten und Entwicklungen zu erheben.
Wichtig erscheint es, die Untersuchungen im intralingualen Vergleich, d.h. für Varietäten, durchzuführen. Auf ein solches Forschungsdesiderat wird auch an anderer Stelle hingewiesen, so etwa bei Ammon (2001: 25), der fordert, „durch konsequente Kontrastierung mit anderen deutschsprachigen Ländern die Südtiroler Spezifik der identifizierten Formen [d.h. der sprachlichen Besonderheiten] zu prüfen." Ein erster Schritt in diese Richtung wurde für die Allgemeinsprache durch das Variantenwörterbuch des Deutschen (Ammon et al. 2004) gesetzt, im Bereich der Rechtssprache etwa durch die Erarbeitung der Terminologischen Datenbank in *bistro*[2]. Umfassende sowie systematische Untersuchungen auf höheren linguistischen Ebenen, für die digitale Ressourcen unabdingbar sind, stehen bislang jedoch aus.
Der vorliegende Beitrag ordnet sich in diesen Kontext ein, und seine Ziele, die im Wesentlichen auf zwei Ebenen anzusiedeln sind, leiten sich daraus ab:
So geht es einmal darum, neue Methoden und Werkzeuge zur Untersuchung von Varietäten besonders auf lexikalischer und syntagmatischer Ebene vorzustellen und damit einen Beitrag zur Methodik in der Variationsforschung zu leisten. Die Verfahren und Instrumente werden in Pilotuntersuchungen anhand eines kleinen Vergleichskorpus von Zeitungstexten eingesetzt und beides, korpuslinguistische Ansätze sowie Datengrundlage, auf ihre Brauchbar-

2 <www.eurac.edu/bistro.htm> (Stand: Jänner 2011)

keit zur Ermittlung sprachlicher Besonderheiten des Südtiroler Deutschen hin analysiert. Zunächst stehen dabei besonders der Ansatz und das Vorgehen bei der Datenerhebung im Mittelpunkt des Interesses; das Ziel ist es, automatisiert eine Vorauswahl der Daten zu erstellen, die anschließend manuell genauer zu untersuchen sind. Damit ist auch die Brücke zum zweiten Untersuchungsschwerpunkt geschlagen, nämlich zur Erhebung und Beschreibung von möglichen Besonderheiten der deutschen Sprache in Südtirol anhand umfangreicher authentischer Datensammlungen, die in Form von elektronischen Korpora vorliegen. Bisherige Ergebnisse in der Varietäten- und Sprachkontaktforschung sollen, wo es möglich ist, untermauert und ergänzt oder revidiert werden. Mit diesem empirischen Ansatz wird ein Beitrag zur Beschreibung der deutschen Sprache Südtirols geleistet.

## 2. Untersuchungen zur geschriebenen deutschen Standardsprache in Südtirol

Untersuchungsschwerpunkte der geschriebenen deutschen Allgemeinsprache in Südtirol bilden sprachliche Besonderheiten bzw. kontaktsprachliche Erscheinungen besonders lexikalischer, aber auch morphologisch-syntaktischer Art. Die Anfänge der Auseinandersetzung mit den Besonderheiten der deutschen Sprache in Südtirol gehen auf die 60er (z.B. Rizzo-Baur 1962) und 70er Jahre (z.B. Riedmann 1972) zurück. Riedmanns Thesen zum Sprachwandel, wie etwa zu den Einflüssen aus dem Italienischen und seine Kritik an sprachlichen Interferenzen, wurden in den darauffolgenden Jahren übernommen (z.B. Kramer 1981, Tyroller 1986), gelten heute jedoch als weitgehend überholt. Die sprachwissenschaftliche Forschung konzentrierte sich weiterhin und im Wesentlichen bis heute auf die Ebene der Lexik (Ein- und Mehrwortebene) und der Grammatik (z.B. Pernstich 1984, Forer/Moser 1988, Lanthaler/Saxalber 1995, Ammon 2001, Ammon et al. 2004; für einen Überblick vgl. Lanthaler 1997), während es kaum Untersuchungen zu Erscheinungen auf syntagmatischer (Kollokationen, idiomatische Wendungen; vgl. Abfalterer 2007: 54) und textueller Ebene (z.B. Riehl 2001; vgl. Spillner 1992: 180) gibt. Wenig untersucht sind außerdem kontaktsprachliche Erscheinungen in übersetzten Texten (z.B. Putzer 1984[3])[4].

Während in den 60/70er und 80er Jahren vielfach „Kontaktphänomene als Beeinträchtigungen des Systems“ (Lanthaler 1995: 209) beschrieben wurden und

3 Analyse übersetzter Texte von KandidatInnen der Zweisprachigkeitsprüfung auf verschiedenen sprachlichen Ebenen

4 Auf Untersuchungen zur gesprochenen Sprache in Südtirol wird im Rahmen dieses Beitrags nicht eingegangen.

die Angst vor einem Sprachverfall vorherrschend war (s. z.B. Riedmann 1972), ging man später von einer Interferenzforschung, die Interferenzen vorwiegend als Anzeichen für einen Verlust wertet, zur Beschreibung von Sonderwortschätzen aus variationslinguistischer Sicht über (vgl. z.B. Egger 2001: 58, Abfalterer 2007: 38, Moser 1990: 20-21).
Die Interpretation bisheriger Forschungsergebnisse weist darauf hin, dass es neben nahe liegenden Übernahmen aus dem Italienischen besonders im Bereich der Amts- und Verwaltungssprache weniger Auffälligkeiten gibt als angenommen (vgl. z.B. Ammon 2001: 25, Pernstich 1984: 112, Masser 1982: 72-73). Dass ein Vorkommen von Italianismen in der deutschen Umgangs- und Hochsprache, v.a. auch in der gesprochenen Sprache, festzustellen ist, steht außer Frage.
Im Zuge der Diskussionen um das Deutsche in Südtirol kommt immer wieder die Frage nach gutem und korrektem Deutsch und in diesem Zusammenhang nach entsprechenden Norminstanzen auf (vgl. z.B. Lanthaler/Saxalber 1995, Ammon 2001, Daniel et al. 2001, Lanthaler 2001). Unsicherheiten in Bezug auf eine korrekte Sprachverwendung scheinen teilweise zu einem Sprachpurismus und zu einer Orientierung an der bundesdeutschen Norm sowie zu einer Ablehnung regionaler Eigenheiten zu führen, was u.a. auf die Bemühungen um den Erhalt der deutschen Hochsprache nach dem Faschismus zurückgeführt wird (vgl. Lanthaler 1995: 293f.). Dies bestätigen beispielsweise Analysen der Berichterstattung zum Thema ‚Sprache in Südtirol' in der auflagenstarken lokalen Tageszeitung ‚Dolomiten' (s. Abfalterer 2007: 37ff.). Neben der wissenschaftlichen Auseinandersetzung mit den Eigenheiten der deutschen Sprache in Südtirol spielen die „'Alltagsrezeption' im Bereich Sprachkritik" (vgl. Abfalterer 2007: 37) und in diesem Zusammenhang die Printmedien eine nicht unbedeutende Rolle. Darüber hinaus sind auch Einzelphänomensammlungen von Sprachinteressierten zu beachten, die einer größeren Öffentlichkeit zur Verfügung gestellt werden (z.B. Daum[5]).
Was die Erhebungsmethoden betrifft, so wurde und wird traditionellerweise mit Belegsammlungen und Exzerpten anhand ausgewählter Textsorten und Texte gearbeitet (z.B. Riedmann 1972, Pernstich 1984, Riehl 2001, Ammon et al. 2004, Abfalterer 2007), wobei anfänglich kaum der systematische intralinguale Vergleich angestrebt wurde. Dieser hingegen rückte im Zuge der Ausarbeitung des Variantenwörterbuchs des Deutschen (Ammon et al. 2004) in den Mittelpunkt des Interesses. Daneben bediente man sich dabei einzelner Rückfragen an Experten und Gewährspersonen und griff auf Wörterbücher und die

5 <http://www.kulturinstitut.org/fileadmin/user_upload/teaserbilder/veranstaltungen/Sprachstelle/SuedtirolerBesonderheiten_ListeDaum.pdf> (Stand: Jänner 2011)

einschlägige Sekundärliteratur zurück (s. dazu z.B. Amonn et al. 2004, Abfalterer 2007). Neuer hingegen ist der Einsatz des Internets als Quelle in der Variationslinguistik (s. dazu z.B. Bickel 2000), etwa bestimmter Internet-Suchmaschinen zur Absicherung und zur Ergänzung herkömmlicher Belegsammlungen im Rahmen der Erarbeitung des Variantenwörterbuchs für das Deutsche (vgl. Bickel 2000: 74, 81, 111). Immer aussichtsreicher für die Variationsforschung wird der Einsatz korpuslinguistischer Methoden. So bildet der so genannte C4-Verbund[6], ein Zusammenschluss von Korpusinitiativen in Deutschland, Österreich, der Schweiz und Südtirol zur Erstellung eines deutschen Vergleichskorpus, eine vielversprechende Grundlage für quantitative vertiefende Studien zu den Varietäten des Standarddeutschen auf verschiedenen sprachlichen Beschreibungsebenen.

## 3. Datengrundlage, methodisches Vorgehen und erste Ergebnisse

### 3.1. Datengrundlage

Die ersten Vergleichsstudien wurden exemplarisch anhand von zwei Zeitungskorpora durchgeführt (vergleichbar mit der Datengrundlage in Heid in diesem Band). Das Textkorpus ‚Dolomiten' (im Folgenden: DOLO) besteht aus rund 66 Mio. Tokens[7] der Südtiroler Tageszeitung ‚Dolomiten' aus den Jahren 1991 bis 2006. Im Textkorpus ‚Frankfurter Rundschau' (im Folgenden: FR) sind rund 40 Mio. Tokens der überregionalen deutschen Tageszeitung ‚Frankfurter Rundschau' aus den Jahren 1992 und 1993 enthalten. Die linguistische Vorverarbeitung beider Korpora umfasst Tokenisierung, Wortart-Annotation, Lemmatisierung und Chunking.[8] Sie wurde automatisch mit dem TreeTagger[9] (Schmid 1994) und dem Chunker YAC (Kermes 2003) durchgeführt. Die Indizierung[10] und Abfrage erfolgte mit der CorpusWorkbench CWB und dem Corpus Query Processor CQP des IMS Stuttgart (Christ 1994, Evert 2005).

6 <http://www.korpus-c4.org> (Stand: Jänner 2011)

7 Als Token wird das „Vorkommen eines Wortes an einer bestimmten Stelle im Korpus" bezeichnet (Lemnitzer/Zinsmeister 2006: 198).

8 Bei der ‚Tokenisierung' wird ein Fließtext in einzelne Einheiten unterteilt (Lemnitzer/Zinsmeister 2006: 64), z.B. werden Wörter von direkt nachfolgenden Satzzeichen getrennt. ‚Annotieren' bezeichnet das Hinzufügen zusätzlicher Informationen zu den Bestandteilen eines Textes (vgl. Lemnitzer/Zinsmeister 2006: 60-88,196). Beim ‚Lemmatisieren' werden die Grundformen (Lemmata) jeder einzelnen Wortform annotiert (Lemnitzer/Zinsmeister 2006: 71). ‚Chunking'-Software fasst zusammengehörige Phrasen innerhalb eines Satzes zusammen; zu unterscheiden von Parsing, bei dem vollständige Syntaxanalysebäume erstellt werden (Lemnitzer/Zinsmeister 2006: 79,198).

9 zur automatischen Annotation von Wortarten (statistisch) und Lemmata (lexikon-basiert)

10 ‚Indizierung' bezieht sich hier auf die Formatierung der Daten für ein bestimmtes Abfragesystem.

Als weitere Ressourcen wurden Ergebnisse der Innsbrucker Forschungen zum Südtiroler Deutschen herangezogen, so die Beiträge zum Variantenwörterbuch des Deutschen (Ammon et al. 2004), die Sammlung von Südtirolismen (Abfalterer 2007) und die CD-ROM ‚Datenbank zum Südtiroler Deutsch'[11]. Zudem wurden Spezialwortschatzlisten, z.B. Namenslisten der Provinz Bozen (Quelle: Landesinstitut für Statistik ASTAT[12]) oder Termlisten der *bistro*-Datenbank[13], verwendet.

Für die Untersuchungen wurden ausschließlich Korpora der Textsorte ‚journalistische Prosa' ausgewählt. Sie können daher nicht als repräsentativ für die deutschen Varietäten in Südtirol und Deutschland angesehen werden. Zudem ist die Datenmenge nicht ausreichend für allgemeingültige Aussagen. Die Korpora sind auch aus inhaltlicher Perspektive nicht direkt und ohne Vorbehalte vergleichbar. Diese Einschränkungen schmälern jedoch keineswegs den Wert der Untersuchungen, die zunächst nicht auf Aspekte wie Repräsentativität und Allgemeingültigkeit hin angelegt, sondern wie eingangs beschrieben auf anderen Ebenen angesiedelt sind. Demnach geht es in erster Linie darum, Ansätze und Instrumente bei der Datenerhebung und -bearbeitung zu präsentieren, die es ermöglichen, eine Vorauswahl von Daten als Anhaltspunkte für detailliertere, manuelle Untersuchungen zu treffen, sowie diese in Pilotuntersuchungen einzusetzen und Ergebnisse vorzustellen.

### 3.2. Einzelstudien

Die Frequenzerhebungen und vergleichenden Untersuchungen sprachlicher Einheiten und Muster gliedern sich in drei Teile[14]:

1) Frequenzerhebung Südtirolismen (Abfalterer 2007) [3.2.1],
2) Untersuchung nicht automatisch lemmatisierter Wörter im DOLO-Korpus (im Folgenden: unknowns) [3.2.2],
3) Extraktion von kontinuierlichen und diskontinuierlichen Kookkurrenzen: [3.2.3]
   a) systematisch: Adjektiv+Substantiv (im Folgenden: Adj+N), Substantiv+Prä-position+Substantiv (im Folgenden: N+Präp+N),
   b) exemplarisch: Substantiv+Verb (im Folgenden: N+V) (Objekt/Subjekt+Prädikat).

---

11 vom Südtiroler Kulturinstitut finanziertes Projekt, durchgeführt am Institut für deutsche Sprache, Literatur und Literaturkritik der Leopold-Franzens-Universität Innsbruck, 2002/2003 (Abfalterer, Moser)

12 <http://www.provinz.bz.it/astat/de> (Stand: Jänner 2011)

13 bistro: Informationssystem für Rechtsterminologie der EURAC <http://www.eurac.edu/bistro> (Stand: Jänner 2011)

14 Ähnliche Untersuchungen zu Rechtstexten sind in Abel/Anstein/Ties (2008) beschrieben.

In den folgenden Abschnitten werden jeweils Methoden und Ergebnisse der drei Studien vorgestellt.

### 3.2.1. Frequenzerhebung Südtirolismen

In Anlehnung an Abfalterer (2007: 167ff) werden lexikalische Besonderheiten des Deutschen in Südtirol für die Untersuchung folgendermaßen unterschieden:

- Primäre Südtirolismen (ausschließlich in Südtirol verwendet):
  - Bestimmte Lemmata werden ausschließlich in Südtirol verwendet und sind in allen ihren Bedeutungen ein Südtirolismus, z.B. *Autobüchlein* (im Folgenden: primS).
  - Bestimmte Lemmata weisen eine spezifische Eigenschaft ausschließlich in Südtirol auf und sind nur in einer Sonderbedeutung ein Südtirolismus, z.B. *Kondominium* (Sonderbedeutung in Südtirol, abgekürzt STIR, im Unterschied zur Verwendung in anderen Zentren des plurizentrischen Deutschen: Österreich, Schweiz, Deutschland, Belgien Liechtenstein, Luxemburg, abgekürzt: A, CH, D, BELG, LIE, LUX) (im Folgenden: primS_SB).
- Sekundäre Südtirolismen (in mindestens einem weiteren Zentrum verwendet):
  - Bestimmte Lemmata sind vollständig konform mit Lemmata in mindestens einem der großen deutschsprachigen Nachbarländer Südtirols, z.B. *Marille* (auch in A) (im Folgenden: sekS).

In der vorliegenden Studie wurden Südtirolismen-Vorkommen (Quelle: Abfalterer 2007: 168-179, 263-268) in den zwei Korpora mit Hilfe von CQP gezählt[15] und die entstandenen Frequenzlisten zunächst mit Perl-Skripten gegenübergestellt (Abb. 1). Dabei war zu erwarten, dass primäre Südtirolismen ohne Sonderbedeutung (primS) in der FR nicht oder sehr selten vorkommen. Einzelne solcher primärer Südtirolismen *mit* Vorkommen in der FR sowie primäre Südtirolismen mit Sonderbedeutung (primS_SB) und sekundäre Südtirolismen (sekS) wurden mit größeren Satzkontexten genauer untersucht.

15 Alle aufgeführten Zahlen bezeichnen absolute oder relative Häufigkeiten (Letztere in Bezug zur Gesamtkorpusgröße) von Vorkommen; statistische Vergleichsmethoden oder Assoziationsmaße wurden in dieser Pilotstudie nicht angewandt. Zudem können unterschiedliche Frequenzdaten auch von verschiedenen Inhalten herrühren oder durch Zitate aus anderen Ländern ‚verfälscht' werden, weswegen eine manuelle Kontextüberprüfung unabdingbar ist. ‚Triviale' Unterschiede wie in Heid in diesem Band beschrieben stellen eine weitere Gruppe dar, z.B. werden tatsächliche Gegebenheiten der betreffenden Länder offensichtlich durch besondere Lexik dargestellt.

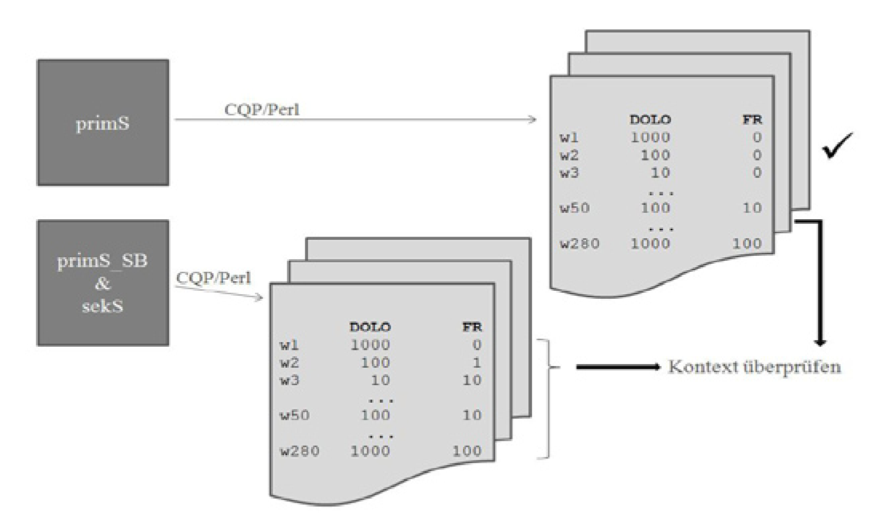



Abb. 1: Frequenzermittlung der Südtirolismen in den zwei Vergleichskorpora

Tabelle 1 stellt einen Auszug aus der Beispielliste der primären Südtirolismen *mit* Vorkommen in der FR dar, die genauer zu betrachten sind.

| | DOLO absolut | DOLO relativ[16] | FR absolut in % | FR relativ[17] in % |
|---|---|---|---|---|
| Abgeordnetenkammer | 801 | 0,00122 | 57 | 0,00142 |
| Schulamtsleiter | 836 | 0,00127 | 48 | 0,00119 |
| SVP | 14566 | 0,00221 | 29 | 0,00721 |
| Regionalregierung | 504 | 0,00766 | 25 | 0,00622 |
| Proporz | 754 | 0,00115 | 21 | 0,00522 |
| Regionalrat | 382 | 0,00581 | 8 | 0,00199 |
| Finanzpolizei | 519 | 0,00789 | 7 | 0,00174 |
| Lyzeum | 205 | 0,00312 | 6 | 0,00149 |
| Vertrauensarzt | 59 | 0,00897 | 5 | 0,00124 |
| Industriellenverband | 587 | 0,00892 | 4 | 0,00099 |
| Schriftleiter | 42 | 0,00638 | 3 | 0,00075 |
| Sanitätseinheit | 394 | 0,00599 | 3 | 0,00075 |
| Carabiniere | 349 | 0,00531 | 3 | 0,00075 |

Tab. 1: Auszug aus der Gegenüberstellung von Frequenzdaten zu primären Südtirolismen ohne Sonderbedeutung, sortiert nach Frequenz in der FR

16 genaue Tokenanzahl DOLO: 65.782.016

17 genaue Tokenanzahl FR: 40.201.367

Als ein erstes Beispiel aus der Gruppe der primären Südtirolismen soll hier exemplarisch das Lemma *Abgeordnetenkammer* genauer beleuchtet werden. Der Eintrag im Wörterbuchteil von Abfalterer (2007), der ähnlich auch im Variantenwörterbuch (Ammon et al. 2004) aussieht, ist in Beispiel (1) dargestellt:

> **(1) Abgeordnetenkammer** die; -, -n <übersetzt aus ital. camera dei deputati; lat. camera>; primärer Südtirolismus; Bed.: ‚erste -> Kammer des italienischen Parlaments'; Bsp.: *Noch in diesem Monat wird die Abgeordnetenkammer in Rom die zweite Lesung des Gesetzes vornehmen* (Dolomiten 16.10.2000, 3); Variation: in A und CH ->Nationalrat, in CH ->Kammer: *grosse Kammer**, in D ->Bundestag, in LIE ->Landtag (Abfalterer 2007)

Die Verwendungen im DOLO-Korpus bestätigen die Bedeutung aus (1), wie die folgenden Beispiele zeigen:

oMit 163 gegen 106 Stimmen billigte der Senat ein neues Anti-Mafia-Gesetz. Die Vorlage geht nun an die <**Abgeordnetenkammer**>, die in der kommenden Woche abstimmen soll.

oDie Neueinteilung der Senatswahlkreise - die in der römischen <**Abgeordnetenkammer**> vor wenigen Tagen neuerlich auf unbestimmte Zeit vertagt wurde – [...]

oDie <**Abgeordnetenkammer**> hat gestern das von der Regierung erlassene Dekret gegen die Wiederverwendung von Schmutzgeldern gutgeheißen.

Zur Kontrolle wurden die Vorkommen in der FR betrachtet, mit dem Schluss, dass sich alle Belegsätze auf Gegebenheiten, in diesem Fall Parlamente, anderer Länder beziehen. Die Begriffe werden als ‚Zitate' verwendet, so z.B.:

oMit großer Mehrheit hat die italienische <**Abgeordnetenkammer**> in Rom das von der Regierung vorgelegte Gesetzespaket zur [...]

oDie <**Abgeordnetenkammer**> des Schweizer Parlaments hat am Mittwoch mit großer Mehrheit für den Vertrag zur Schaffung eines Europäischen Wirtschaftsraums (EWR) gestimmt.

oMit großer Mehrheit hat die belgische <**Abgeordnetenkammer**> am Wochenende dem Vertrag von Maastricht für eine Europäische Union mit gemeinsamer Währung zugestimmt.

Als ein weiteres Beispiel wird ein Lemma untersucht, das nur in einer Sonderbedeutung ein Südtirolismus ist, u.z. das Lemma *Konsortium*. In diesem Fall muss in der Korpusarbeit der jeweilige Kontext genau betrachtet werden, da – aufgrund der identischen Wortform – nicht automatisch klar ist, dass semantische Unterschiede bestehen.
Ein *Konsortium* ist nach Abfalterer (2007) in Beispiel (2) in der Südtiroler Sonderbedeutung eine dauerhafte Gemeinschaft, beschränkt auf die öffentliche Verwaltung und Landwirtschaft, im Gegensatz zu einem vorübergehenden Zusammenschluss wie in der gemeindeutschen Bedeutung nach (3):

**(2)** **Konsortium** das; -s, …ien; <aus ital. consorzio; lat. consortium ‚Teilnahme'>; primärer Südtirolismus in der Sonderbed.: ‚dauerhafte Arbeits- und Interessengemeinschaft in der öffentlichen Verwaltung oder in der Landwirtschaft'; Bsp.: *Die einheitliche Verwaltung des Parks soll durch ein Konsortium zwischen dem Staat… und den Autonomen Provinzen Bozen und Trient erfolgen* (Südtirols Autonomie, 226) – Die Bedeutung ‚vorübergehender Zusammenschluss von Unternehmen, z.B. Banken' ist gemeindt. […] (Abfalterer 2007)

**(3)** **Konsortium** […] (Wirtsch.): vorübergehender Zusammenschluss von Unternehmen, bes. Banken, zur gemeinsamen Durchführung eines größeren Geschäfts: ein internationales K. von elf Großbanken; ein K. bilden, gründen (Duden Universalwörterbuch 2001)

Die Südtiroler Sonderbedeutung mit den semantischen Merkmalen nach (2) bestätigen viele Beispiele aus der DOLO, so z.B.:

o Der Verkehrsverbund - VVB - ist ein <**Konsortium**> zwischen den Gemeinden Bozen, Meran und Leifers und führt den öffentlichen Personentransport durch […]

o Sie würden sich bereit erklären, die Anpassung des Pachtzinses zu akzeptieren. Dann jedoch müßte das <**Konsortium**>, so die Bauern, die Meliorierung und Kultivierung des Grundes entsprechend entschädigen […]

o […] ein Agrar- und Lebensmittelgeschäft gegründet. 1956 wurde die Niederlassung, die bei den Bauern besser als „<**Konsortium**>" bekannt […]

Weitere Belege aus der DOLO zeigen aber auch, dass dem Konsortium in Südtirol eine weiter gefasste Bedeutung zukommt:

- Zwei afrikanische Köchinnen werden ein leckeres Abendessen zubereiten. Die Südtiroler Weltläden sowie das <**Konsortium**> „CTM Altromercato" werden sich und ihre Produkte vorstellen […]
- Grund dafür war vor allem die Rechtsform <**Konsortium**>: Sie erschien den Mitgliedern nicht mehr zeitgemäß, da damit vor allem in steuerrechtlicher Hinsicht hohe Kosten verbunden waren […]

Als Lehnwort[18] aus dem Italienischen scheint das *Konsortium* alle wesentlichen Merkmale der Originalsprache zu übernehmen, wie der Wörterbucheintrag[19] (4) zeigt

**(4)** **con|sòr|zio**, s.m. […]
2 TS econ., fin., associazione di più soggetti, costituita per coordinare e regolamentare le attività comuni, spec. economiche o finanziarie: c. di piccole e medie imprese, c. industriale, provinciale, c. tra produttori vinicoli […] Polirematiche: consorzio agrario loc.s.m. […]

Ausgehend von diesen Belegen und Untersuchungen kann die Beschreibung der Südtiroler Sonderbedeutung in den erwähnten Quellen ergänzt bzw. präzisiert werden: *Konsortium* ist demnach ein Lehnwort aus dem Italienischen mit den Merkmalen ‚Verbund/Zusammenschluss', ‚privat/öffentlich', zur ‚Koordination/Befriedigung gemeinsamer (besonders wirtschaftlich-finanzieller) Interessen', wobei die zeitliche Dauer unwesentlich ist und keine Beschränkung auf Bereiche wie öffentliche Verwaltung oder Landwirtschaft besteht.

In einer zusätzlichen ähnlichen Studie im Rahmen der Arbeiten an *Vis-À-Vis*[20] (s. auch Anstein 2009), einem System zum Vergleich von Varietäten anhand von Korpora, wurden Wortlisten der beiden Korpora erstellt und nach Frequenzen sortiert. Tabelle 2 zeigt einige Beispiele mit hoher Frequenz in der DOLO und geringerer Frequenz in der FR.

---

18 nach der Terminologie Beltz (1974), auch in Abfalterer (2007:155): Lehnwörter = Übernahmen aus dem Ital. <(aus) ital.>; Lehnbildungen = Übersetzungen < übersetzt aus ital.>

19 Online-Version des italienischen Wörterbuchs De Mauro – Paravia: <http://www.demauroparavia.it/26415> (Stand: März 2008; 2011: geänderte Zugriffsmodalitäten)

20 <http://www.korpus-suedtirol.it/korpus-suedtirol_II_de.htm> (Stand: Jänner 2011)

| | DOLO absolut | DOLO relativ in % | FR absolut | FR relativ in % |
|---|---|---|---|---|
| hingegen | 12782 | 0,001943 | 2358 | 0,00587 |
| Rang | 10940 | 0,001663 | 2876 | 0,00715 |
| Start | 10891 | 0,001656 | 2432 | 0,00605 |
| Vorjahr | 10620 | 0,001614 | 1813 | 0,00451 |
| Sprache | 9838 | 0,001496 | 3003 | 0,00747 |
| organisieren | 9724 | 0,001478 | 3569 | 0,00888 |
| Bauer | 9478 | 0,001441 | 2518 | 0,00626 |
| erzielen | 9400 | 0,001429 | 2853 | 0,00710 |
| erfolgen | 9296 | 0,001413 | 1808 | 0,00450 |
| Direktor | 9272 | 0,001410 | 1239 | 0,00308 |
| betreffen | 9157 | 0,001392 | 3079 | 0,00766 |
| Kreuz | 9101 | 0,001384 | 1970 | 0,00490 |
| Beitrag | 9099 | 0,001383 | 3125 | 0,00777 |
| Lauf | 8336 | 0,001267 | 2284 | 0,00568 |
| somit | 8241 | 0,001253 | 1762 | 0,00438 |
| morgig | 7847 | 0,001193 | 785 | 0,00195 |
| gestrig | 7839 | 0,001192 | 484 | 0,00120 |
| Leitung | 7769 | 0,001181 | 2664 | 0,00663 |
| zufrieden | 7601 | 0,001155 | 2912 | 0,00724 |
| Wettbewerb | 7512 | 0,001142 | 2361 | 0,00587 |
| Papst | 7463 | 0,001135 | 471 | 0,00117 |
| (sich) klassieren | 513 | 0,007798 | 10 | 0,00249 |

Tab. 2: Auszug aus der Gegenüberstellung von Wortlisten der beiden Vergleichskorpora, sortiert nach Frequenz in der DOLO

Solche Hoch- vs. Niedrigfrequenz-Gegenüberstellungen können Hinweise auf bisher nicht erfasste lexikalische Besonderheiten liefern. Für Phänomene, die eine jeweils unverhältnismäßig hohe bzw. niedrige Frequenz in den Vergleichskorpora aufweisen, kann in Detailkontextstudien das Vorhandensein von Sonderbedeutungen überprüft werden.

### 3.2.2. Untersuchung der unknowns

In der zweiten Teilstudie wurde die beim Lemmatisieren mit dem TreeTagger entstandene *unknowns*-Liste aus der DOLO mit ca. 832000 Types als eine Quelle für neue Südtirolismen oder anderen Spezialwortschatz genauer be-

trachtet. Für die in die Liste aufgenommenen Wörter wurde in dem für das Bundesdeutsche entwickelten Tagger (Schmid 1994) kein Lexikoneintrag gefunden, was auf möglicherweise südtirolspezifischen Wortschatz schließen lässt. Nach dem mithilfe von Perl-Skripten automatischen Filtern der *unknowns*-Liste mit bestehenden Spezialwortschatzlisten (bekannte Varianten, Namen, etc.; siehe zu den Quellen auch Abschnitt 3.1) ergibt sich eine Wortliste für neue Kandidaten, die manuell auszuwerten ist. Die Filterlisten können jeweils semiautomatisch mit den neuen Kandidaten vervollständigt werden (zum Ablauf s. Abb. 2).

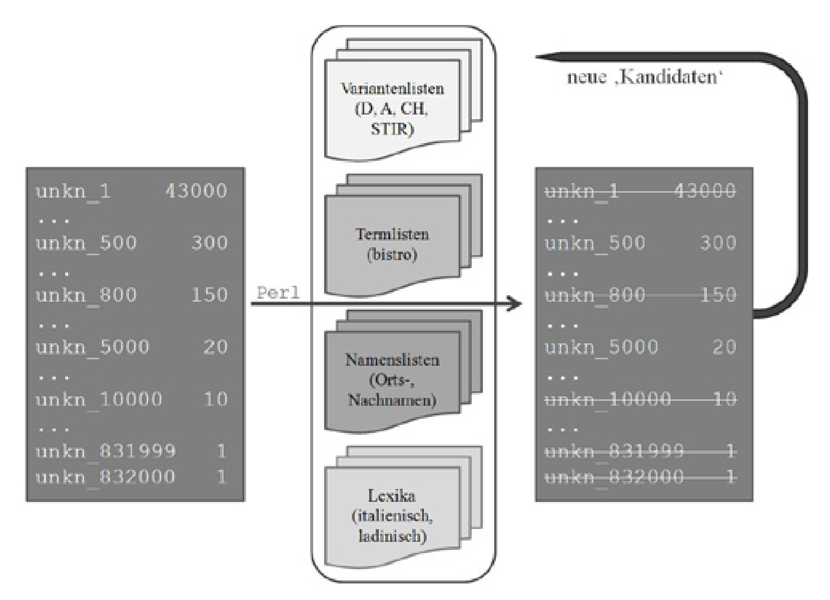



Abb. 2: Ablauf der *unknowns*-Studie

Tabelle 3 zeigt einen Auszug aus der sich ergebenden Liste der *unknowns*. Einheiten, die keine automatische Klassifikation erhielten, sind mit einer leeren eckigen Klammer gekennzeichnet. Diese wurden manuell in vorgegebene Klassen eingeteilt und ggf. lemmatisiert, um sie für weiteren Gebrauch mit mehr Information anzureichern.

| | Frequenz DOLO absolut | automatische Klassifizierung | manuelle Klassifizierung | ggf. Lemma |
|---|---|---|---|---|
| Tschengls | 447 | [] | Ortsname deutsch | |
| Bressanone | 122 | [] | Ortsname italienisch | |
| Mayr | 7094 | [] | Nachname | |
| SVP | 14566 | [] | Parteiname Abkürzung | |
| Zett | 3600 | [] | Produktname Zeitung | |
| Vinschgaus | 596 | [] | Flurname | Vinschgau |
| weiters | 2778 | [Austriazismus] | auch STIR | |
| ober$_{\text{Präp}}$ | 231 | [Austriazismus] | auch STIR | |
| Stadel | 624 | [Austriazismus] | auch STIR | |
| ehestens | 123 | [] | Südtirolismus | |

Tab. 3: Südtirolismenkandidaten aus den *unknowns*

Die als Austriazismen geltenden lexikalischen Einheiten weiters, ober als Präposition und Stadel sowie das Adverb ehestens sind neue Kandidaten für Südtirolismen: Für weiters mit 2778 Belegen in der DOLO, ober mit 231 und Stadel mit 624, die im Variantenwörterbuch bei Ammon et al. (2004) nicht den Zusatz STIR aufweisen (s. dort: weiters: A, LIE, Adv.: […]; ober: A, Präp. mit Dat. […]; Stadel: A, CH-ost, D-südost) und bei Abfalterer (2007) als Einträge gänzlich fehlen, sowie für ehestens mit 123 Belegen in der DOLO in der Bedeutung baldmöglichst, das in beiden Wörterbüchern (Ammon et al. 2004 und Abfalterer 2007) nicht erwähnt ist, können so entsprechende Ergänzungen für eventuelle Neuauflagen vorgeschlagen werden. Die folgenden Wörterbuch- und Korpusbelege helfen, die genannten Beispiele zu veranschaulichen.

**(5)** **weiters** A LIE Adv.: ➔ WEITER CH D >ferner, weiterhin, außerdem<: [...] - Vgl. durchwegs, öfters (Variantenwörterbuch 2004)

**(6)** **weiters** <Adv.> (österr.): *weiterhin* (3) (Duden Universalwörterbuch 2001)

`o<`**`Weiters`**`> genehmigte der Rat die Veräußerung eines Gemeindegrundstückes von rund 500 Quadratmetern. (DOLO)`

**(7)** **ober** A Präp. mit Dat.: ➔ OB CH >oberhalb von; über<: […] - In D (landsch.) veraltet (Variantenwörterbuch 2004; kein Eintrag im Duden Universalwörterbuch 2001)

o Über die Mauerrisse sind Rauchgase direkt in die Kammer <ober> der Stube gedrungen, in der H. lag. (DOLO)

**(8)** **Stadel** A CH-ost D-südost der; -s, -n: ➔ SCHEUNE A-west CH D-nord/mittel, ➔ SCHOBER A-west CH D-mittelost/südwest, ➔ SCHEUER D-südwest >Teil des Bauernhauses oder freistehendes Gebäude, in dem Heu und Stroh, oft auch Gerätschaften, aufbewahrt werden<: […] - Dazu: Heustadel, Holzstadel (Variantenwörterbuch 2004)

**(9)** **Stadel**, der; -s, - u. (schweiz.:) Städel, (österr. auch:) Stadeln […] (süddt., österr., schweiz.): *Heustadel* (Duden Universalwörterbuch 2001)

**(10)** **Heustadel**, der (südd., österr., schweiz.): *Scheune zum Aufbewahren von Heu* (Duden Universalwörterbuch 2001)

o Die Lüsener Feuerwehr konnte weitere Brandherde im <**Stadel**> und in der Sennhütte vor Ausbruch des Feuers entschärfen. (DOLO)

**(11)** **ehestens** <Adv.>: 1. *frühestens*: e. [am] Dienstag [...] 2. (österr., sonst veraltet) *baldmöglichst; so schnell, bald wie möglich* (Duden Universalwörterbuch 2001)

o Ebenso müßte die Säuberung der Straßen vom Streusand <**ehestens**> vorgenommen werden. (DOLO)

o Im Wohnbausektor will die Verwaltung die Erweiterungszone „Gebach" <**ehestens**> verwirklichen. (DOLO)

Zur weiteren Erleichterung der manuellen Feinarbeit sollten künftig bereits bekannte und erfasste Südtirolismen in eine morphologische Analysekomponente eingebaut werden, um Kompositabildungen mit einem Südtirolismus als Bestandteil mithilfe computerlinguistischer Werkzeuge zu zerlegen. So können *unknowns*-Einträge der Art *Vizequästor* oder *Ökonomatskasse*, bei denen jeweils nur die Konstituenten *Quästor* und *Ökonomat* Südtirolismen sind, automatisch herausgefiltert werden.

### 3.2.3. Extraktion von kontinuierlichen und diskontinuierlichen Kookkurrenzen

In der letzten der drei Teilstudien werden Ausführungen und Ergebnisse vorgestellt, in denen Listen kontinuierlicher und diskontinuierlicher Kookkurrenzen aus beiden Korpora mithilfe von CQP und Perl nach bestimmten Mustern von Wortarten extrahiert und nach Häufigkeit sortiert wurden (s. Abb. 3). Als ‚Kookkurrenz' wird hier das gemeinsame Auftreten zweier lexikalischer Einheiten bezeichnet; ‚kontinuierlich' beschreibt Elemente, die direkt aufeinander folgen, ‚diskontinuierlich' dagegen solche, die innerhalb eines Satzes vorkommen und durch funktionale Relationen, z.B. Subjekt-Prädikat, verbunden sind (s. dazu auch Abel/Anstein/Ties 2008).

Kontinuierliche Kookkurrenzen wie Adj+N oder N+Präp+N wurden systematisch behandelt, syntaktische Relationen wie Objekt/Subjekt+Prädikat exemplarisch, da bei Letzteren mit den bisherigen Methoden keine systematischen Unterschiede zwischen den zwei Korpora festgestellt werden konnten. Für N-V-Kookkurrenzen decken die Anfragen verschiedene Satzmuster (Verbzweit- und Verbletztsätze) sowie verschiedene Tempora (zusammengesetzte und einfache) ab; mehrere CQP-Anfragen führen also zu einer Kookkurrenzliste pro Muster. Die Ergebnisse aus den zwei Korpora wurden verglichen, um auch hier durch die genauere Untersuchung von seltenen Fällen in der FR Besonderheiten des Südtiroler Deutschen aufzudecken.

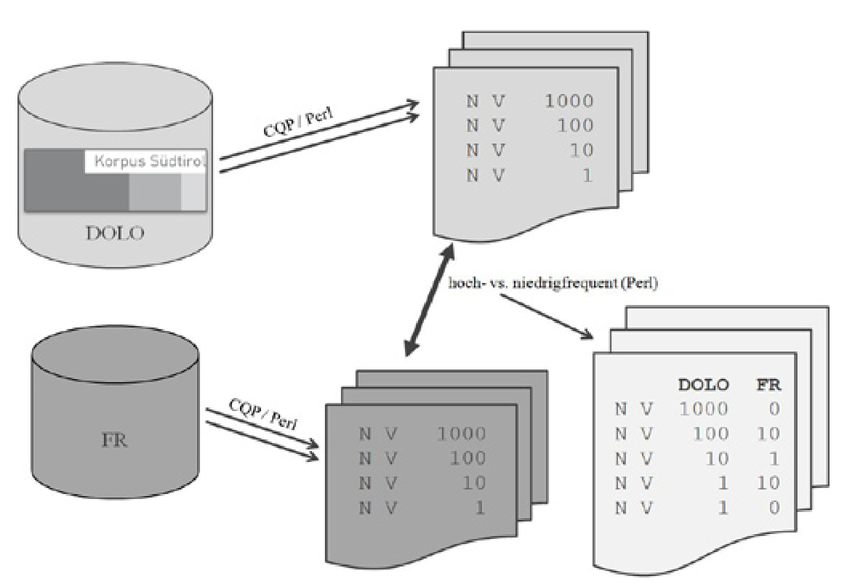



Abb. 3: Arbeitsablauf der Kookkurrenzstudie

Für Adj+N- sowie (N+)Präp+N-Kookkurrenzen werden im Folgenden verschiedene Kontextbeispiele aufgeführt:

a) *allgemein(e) Klasse* (DOLO: 522 vs. FR: 0)

o Auf dem Programm stehen insgesamt sechs Läufe in den Altersklassen Jugend (U-17), Junioren (U-19) und <**allgemeine Klasse**> (jeweils Männer und Frauen). (DOLO)

o Es wird dabei in zwei Kategorien gespielt: <**allgemeine Klasse**> und Altherren. (DOLO)

Die *allgemeine Klasse*, die in der DOLO im Zusammenhang mit verschiedenen Sportarten vorkommt und diese Gruppe von anderen, wie z.B. ‚Jugendlichen' oder ‚Altherren', unterscheidet, wird in der FR in vergleichbaren Kontexten nicht verwendet.

b) *weiß(er) Stimmzettel* (DOLO: 81 vs. FR: 2)

o Mit unserem Programm wollen wir auch jene ansprechen, die das letzte Mal einen <**weißen Stimmzettel**> abgegeben haben, weil sie sich von keiner Partei angesprochen fühlten, […] (DOLO)

o Danach folgen auf den <**weißen Stimmzetteln**> für die Wahl der Gemeindeparlamente und den gelben für die Ortsbeiräte die örtlichen Wählergemeinschaften […] (FR)

In Südtirol bezeichnet ein *weißer Stimmzettel* einen nicht ausgefüllten, in Übereinstimmung mit der italienischen Bezeichnung *scheda bianca*, wohingegen sich das Adjektiv in Deutschland auf die Farbe des Papiers bezieht.

c) *innerhalb* <Monat>/<Wochentag> (DOLO: 420 vs. FR: 0)

o <**Innerhalb März**> können die Fragebögen […] abgeholt und ausgefüllt wieder abgegeben werden. (DOLO)

o Wer sich nicht <**innerhalb Freitag**> die Gutschrift ausstellen lässt, verliert jeden Anspruch auf etwaige Geldrückgabe […] (DOLO)

Auch die Verwendung von *innerhalb* mit Zeitpunkten (d.h. Angabe von Wochentagen, Monaten, Uhrzeiten) in Südtirol anstelle von *bis* ist durch das italienische *entro* motiviert. Gemeindeutsch wird *innerhalb* hingegen im Zusammenhang mit Zeiträumen (inkl. Fristen bestimmten Ausmaßes) verwendet,

wie der Wörterbucheintrag (12) zeigt und was sich darüber hinaus auch im umfangreichen Korpus des DWDS[21] bestätigen lässt.

**(12)** **innerhalb** […] I. <Präp. Mit Gen.> a) in einem bestimmten unbegrenzten Raum; im Inneren, im Bereich […]; b) in einem bestimmten Zeitraum, im Verlauf von, während, binnen: i. der Ferien; i. einer Woche; <mit Dativ, wenn der Genitiv formal nicht zu erkennen ist:> i. fünf Monaten. II. <Adv. In Verbindung mit „von“> a) im Inneren, im Bereich: i. von Berlin […]; b) im Verlauf von, während, binnen: i. von drei Jahren (Duden Universalwörterbuch 2001)

d) *Achtung auf* (N) (DOLO: 11 vs. FR: 0)

o Man findet sie oft auf Mineralienbörsen billig zum Kauf angeboten (<**Achtung auf**> Nachahmungen. (DOLO)

o Folgende Dokumente sind für die Ausstellung eines Passes nötig: Zwei Fotos ([…] Gesicht muss gut sichtbar sein – <**Achtung auf**> Stirnfransen!), […] (DOLO)

Der Gebrauch der Präposition *auf* nach *Achtung* scheint ebenfalls durch die Kombination *attenzione a* im Italienischen begründet zu sein. Gemeindeutsch wird *Achtung* nur in der Bedeutung von *Hoch-/Wertschätzung* mit einer Präposition wie in (13) verwendet. Beispiele im DWDS-Korpus zusätzlich zur FR belegen dies ebenfalls und zeigen zudem, dass *Achtung* (im Zeitraum 1970–2000) in keinem der Fälle mit der Präposition *auf*, sondern vorwiegend mit *vor*, gefolgt von *für* und *gegenüber* verwendet wird.

**(13)** **Achtung**, die; - […]: 1. Hoch-, Wertschätzung, Respekt: […] vor jmdm., etw. A. haben; aus A. vor seinen Eltern […]; 2. Als Ruf od. Aufschrift, um zur Vorsicht oder Aufmerksamkeit zu mahnen: A.!; A., Stufe!; A., Hochspannung!; A., Aufnahme!; 3. […] (Duden Universalwörterbuch)

e) *(erst) in einem zweiten Moment* (DOLO: 47 vs. FR: 0)

o Unklar ist, ob die „Europasteuer“ ins Finanzgesetz für 1997 aufgenommen wird oder <**erst in einem zweiten Moment**> greift. (DOLO)

21 Korpus des DWDS (Digitales Wörterbuch der Deutschen Sprache): <http://www.dwds.de> (Stand: Jänner 2011)

o Worin die Gegenleistung besteht, legen wir aber <**erst in einem zweiten Moment**> mit Durchführungsverordnung fest, [...] (DOLO)

Das Vorkommen der Wendung *in einem zweiten Moment* in der DOLO, vergleichbar etwa mit gemeindeutsch *zu einem späteren Zeitpunkt* (so belegt in der FR und z.B. auch im DWDS-Korpus), scheint ebenfalls durch Interferenz aus dem Italienischen (*in un secondo momento*) erklärbar zu sein.

Zuletzt wurden einige Detailstudien zu einzelnen auffälligen N+V-Kookkurrenzen durchgeführt, die einer weiteren systematischen Untersuchung bedürfen, u.a. zum Fixiertheitsgrad, zur möglichen Anfälligkeit bestimmter morphosyntaktischer Merkmale für Veränderungen, etc. Da die Phänomene in den beiden untersuchten Korpora sehr wenige Belege aufweisen, kann es sich um Zufallsfunde bzw. Tippfehler und möglicherweise nicht um südtirolspezifische Eigenheiten handeln. Zur Absicherung sind zusätzliche Analysen anhand größerer Datenmengen und weiterer Quellen nötig.
Zwei Fälle sollen im Folgenden exemplarisch beleuchtet werden:

f) *Geschäft machen*

o Eines haben sie gemeinsam. Sie wollen ihren Umsatz steigern, ganz einfach <**Geschäft machen**>. (DOLO)

o [...] daß die Presse mit zwei Männern <**Geschäfte macht**> , von denen der eine [...] (FR)

o Wir hatten das Gefühl , mit den Behinderten soll nur ein <**Geschäft gemacht**> werden [...] (FR)

Auffallend ist der Beleg *Geschäft machen* in der DOLO, da man gemeindeutsch die Kombination *Geschäft (Sg.)* + *machen*, in der *Geschäft* die Bedeutung *Profit/Gewinn* trägt, vorwiegend in Verbindung mit einem vorangehenden Artikelwort (*der, ein, kein, mein*) (mit eventuell dazwischen stehenden Attributen) findet. Ansonsten ist die Form *Geschäfte (Pl.)* + *machen* üblich, wobei *Geschäft* im Sinne von *auf Gewinn abzielende Unternehmung* verwendet wird. Dafür sprechen neben dem Wörterbucheintrag in (14) auch eine Vielzahl von Belegen nicht nur in der FR, sondern auch in der Kookkurrenzdatenbank des IDS sowie im Korpus und in den Wortprofilen des DWDS[22].

22 Kookkurrenzdatenbank des IDS: <http://corpora.ids-mannheim.de/ccdb> (Stand: Jänner 2011); Korpus und Wortprofile des DWDS: <http://www.dwds.de> und <http://odo.dwds.de/wortprofil> (Stand: Jänner 2011; Letzteres mit passwortgeschütztem Zugang)

**(14) Geschäft**, das; -[e]s, -e […]: 1.a) auf Gewinn abzielende [kaufmännische] Unternehmung, [kaufmännische] Transaktion; Handel: die –e gehen gut; […] mit jmdm. –e machen […]; b) <o. Pl.> die kaufmännische Transaktionen; Verkauf, Absatz […]; c) <o. Pl.> Gewinn [aus einer kaufmännischen Unternehmung], Profit: diese Unternehmung war für uns [k]ein G […]; ein G. wittern; […] er hat damit ein [glänzendes] G. gemacht ([sehr] viel damit verdient); […] (Duden Universalwörterbuch 2001)

g) *Leistung bringen*

oIch gehe einmal davon aus, daß die Torleute eine <**reife Leistung bringen**>. (DOLO)

Im Beispiel *eine reife Leistung bringen* in der DOLO, das so weder in der FR noch in anderen Korpora wie dem DWDS-Korpus, der Kookkurrenzdatenbank des IDS oder im Wörterbucheintrag (15) vorzufinden ist, scheinen verschiedene feste Wendungen, nämlich *reif* + *Leistung* und *Leistung* + *bringen*, vermischt worden zu sein. Vertiefende Analysen der Belege in der IDS-Datenbank zeigen, dass *reif* + *Leistung* einerseits allein stehend vielfach als Ausruf oder in Verbindung mit dem Verb *sein* als Prädikatsnomen vorkommt sowie andererseits als direktes Objekt besonders häufig in Verbindung mit dem Verb *zeigen*, daneben auch mit *bieten*, *vorweisen* etc. (nicht mit *bringen*) verwendet wird.

**(15) Leistung**, die; -, -en […]: 1. <o. Pl.> *das Leisten* (1). 2. a) *etw. Geleistetes; geleistete körperliche, geistige Arbeit; unternommene Anstrengung u. das erzielte Ergebnis*: eine hervorragende, gute, schlechte, mangelhafte, schwache L.[…]; L. bringen (ugs.; *etw. leisten*) […]; reife L.! (ugs.; *großartig!*) […] (Duden Universalwörterbuch 2001)

## 4. Fazit und Ausblick

Die vorgestellten Studien geben Hinweise darauf, dass mit dem für die Varietätenuntersuchungen gewählten semiautomatischen Ansatz in kürzerer Zeit größere Datenmengen bewältigt werden können, als dies rein manuell möglich ist. Die mithilfe computerlinguistischer Werkzeuge geleistete Vorauswahl der Daten erlaubt es Linguisten, Lexikographen oder anderen Sprachexperten, sich bei der manuell zu leistenden Arbeit vorwiegend auf die linguistische

Feinarbeit zu konzentrieren, z.B. auf semantische und morpho-syntaktische Detailuntersuchungen von Einzellexemen oder Kookkurrenzen.
Die beispielhaft vorgestellten Detailstudien konnten bisherige Beschreibungen von Südtiroler Besonderheiten sowohl untermauern als auch ergänzen. Zugleich lieferten sie wichtige Anhaltspunkte für die nächsten Arbeitsschritte, die im quantitativen und qualitativen Ausbau der Korpora, in der Verbesserung der Qualität der Datenauszeichnung und der spezifischen Weiterentwicklung von Werkzeugen einerseits, sowie in der Konzentration der linguistischen Untersuchung auf bestimmte Phänomene andererseits liegen.
Dazu werden die Datengrundlagen weiter angereichert, wodurch ihre Vergleichbarkeit zusätzlich erhöht wird. Zudem ermöglicht der Ausbau der bereits vorhandenen Daten zu einem Südtirol-Referenzkorpus im Rahmen der C4-Initiative Vergleiche mit österreichischen und Schweizer Korpora.
Die Instrumente zum automatisierten Varietätenvergleich anhand von Korpora werden weiterentwickelt und in einem komplexen, aber einfach zu bedienenden System namens *Vis-À-Vis* zusammengefasst. Auch die Annotationswerkzeuge werden angepasst, denn die Qualität der Ergebnisse von automatisiertem Korpusvergleich hängt doch wesentlich von der Qualität der linguistischen Annotation ab. Da Werkzeuge meist für die ‚Hauptvarietäten' entwickelt wurden, sind besonders Abweichungen anfällig für Auszeichnungsfehler.
Die quantitativen und qualitativen Phänomenuntersuchungen auch im Vergleich mit anderen Varietätenkorpora werden fortgeführt, zunächst u.a. zu folgenden Schwerpunkten:

- Südtirolismen und deren Kollokatoren,
- Systematisierung und Aufarbeitung vorhandener Studien und Phänomensammlungen (z.B. unbearbeitete Einträge der ‚Datenbank zum Südtiroler Deutsch' (Dialektwörter, Einzelfunde, Phraseologismen, Komposita, Eigennamen, etc.),
- synthetische vs. analytische Konstruktionen (z.B. *-ministerium* vs. *Ministerium für ...*),
- Kollokationen/Phraseologismen (z.B. Fixiertheitsgrad, lexikalische und morpho-syntaktische Besonderheiten, etc.),
- Ursachenklärung (z.B. Kontaktphänomene, Sprachwandel).

In einem weiteren Schritt können die Erkenntnisse über Varietäten wiederum für eine korrektere Annotation von Korpora und die Anpassung von Annotationswerkzeugen verwendet werden. So werden zunehmend bessere Grundlagen für weitere Korpus(vergleichs)studien geschaffen, die eine wichtige Voraussetzung z.B. für die lexikographische Arbeit darstellen.

Die Ergebnisse der korpuslinguistischen Arbeit in Südtirol kommen zudem nicht nur der sprachwissenschaftlichen und computerlinguistischen Forschung zugute, sondern können zusätzlich sowohl im Bereich der Lehrerausbildung als auch der Sprachdidaktik eingesetzt und darüber hinaus als wichtige Grundlage für bildungspolitische Entscheidungen genutzt werden.

## Literaturverzeichnis

Abel, Andrea / Anstein, Stefanie / Ties, Isabella (2008): *Ansätze einer intralingualen kontrastiven Korpuslinguistik – aufgezeigt am Beispiel administrativer Rechtstexte aus Deutschland, Österreich und Südtirol.* In: Heller, Dorothee (Hg.): Formulierungsmuster in deutscher und italienischer Fachkommunikation. Intra- und interlinguale Perspektiven; Linguistic Insights 89. Bern: Peter Lang, 243-270.

Abfalterer, Heidemaria (2007): *Der Südtiroler Sonderwortschatz aus plurizentrischer Sicht. Lexikalisch-semantische Besonderheiten im Standarddeutsch Südtirols.* Innsbruck: University Press.

Ammon, Ulrich (1995a): *Vorschläge zur Typologie nationaler Zentren und nationaler Varianten bei plurinationalen Sprachen – am Beispiel des Deutschen.* In: Muhr, Rudolf / Schrodt, Richard / Wiesinger, Peter (Hgg.): Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachliche Aspekte einer nationalen Variante des Deutschen. Wien: Hölder-Pichler-Tempsky, 111-120.

Ammon, Ulrich (1995b): Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten. Berlin-New York: de Gruyter.

Ammon, Ulrich (2001): *Die Plurizentrizität des Deutschen, oder: Wer sagt, was gutes Deutsch ist?* In: Egger, Kurt / Lanthaler, Franz (Hgg.): Die deutsche Sprache in Südtirol: Einheitssprache oder regionale Vielfalt. Wien – Bozen: Folio, 11-26.

Ammon, Ulrich / Bickel, Hans / Ebner, Jakob / Esterhammer, Ruth / Gasser, Markus / Hofer, Lorenz / Kellermeier-Rehbein, Birte / Löffler, Heinrich / Mangott, Doris / Moser, Hans / Schläpfer, Robert / Schloßmacher, Michael / Schmidlin, Regula / Vallaster, Günter (2004): *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol.* Berlin-New York: de Gruyter.

Anstein, Stefanie (2009): *Vis-À-Vis - a System for the Comparison of Linguistic Varieties on the Basis of Corpora.* In: Lyding, Verena (Hg.): Proceedings of LULCL II 2008. Bozen: EURAC 59-64.

Beltz, Werner (1974[3], neub.) [1959]: *Lehnwörter und Lerhnprägungen im Vor- und Frühdeutschen*. In: Maurer, Friedrich / Stroh, Friedrich (Hgg.): Deutsche Wortgeschichte. Bd. 1. Berlin-New York: de Gruyter, 135-163.

Bickel, Hans (2000): *Das Internet als Quelle für die Variationslinguistik.* In: Häcki Buhofer, Annelies (Hg.): Vom Umgang mit sprachlicher Variation: Soziolinguistik, Dia-

lektologie, Methoden und Wissenschaftsgeschichte. Tübingen – Basel: Francke (= Basler Studien zur deutschen Sprache und Literatur; Bd. 80), 111-124.

Christ, Oliver (1994): *A Modular and Flexible Architecture for an Integrated Corpus Query System.* In: Proceedings of COMPLEX 1994, Ungarn, 23-32.

Daniel, Erich / Egger, Kurt / Lanthaler, Franz (2001): *Sprachnormautoritäten in Südtirol*. In: Egger, Kurt / Lanthaler, Franz (Hgg.): Die deutsche Sprache in Südtirol: Einheitssprache oder regionale Vielfalt. Wien – Bozen: Folio, 208-231.

Duden – Deutsches Universalwörterbuch (2001[4]). Hg. von der Dudenredaktion. Mannheim-Leipzig-Wien-Zürich: Dudenverlag.

Egger, Kurt (2001): Sprachlandschaft im Wandel. Südtirol auf dem Weg zur Mehrsprachigkeit. Bozen: Athesia.

Evert, Stefan (2005): *The CQP query language tutorial*. Universität Stuttgart. <http://cwb.sourceforge.net> (Stand: Jänner 2011)

Forer, Rosa / Moser, Hans (1988): *Beobachtungen zum westösterreichischen Sonderwortschatz*. In: Wiesinger, Peter (Hg.): Das österreichische Deutsch. Wien – Köln – Graz: Böhlau, 189-209.

Kramer, Johannes (1981): *Deutsch und Italienisch in Südtirol.* Heidelberg (= Reihe Siegen. Beiträge zur Literatur- und Sprachwissenschaft, Bd. 23).

Kermes, Hannah (2003). *Off-line (and On-line) Text Analysis for Computational Lexicography*. Dissertation: Universität Stuttgart.

Lanthaler, Franz / Saxalber, Annemarie (1995): *Die deutsche Standardsprache in Südtirol*. In: Muhr, Rudolf / Schrodt, Richard / Wiesinger, Peter (Hgg.): Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachliche Aspekte einer nationalen Variante des Deutschen.Wien: Hölder-Pichler-Tempsky, 289-305.

Lanthaler, Franz (1997): *Varietäten des Deutschen in Südtirol*. In: Stickel, Gerhard (Hg.): Varietäten des Deutschen: Regional- und Umgangssprachen. Berlin-New York: de Gryter, 364-383.

Lanthaler, Franz (2001): *Zwischenregister der deutschen Sprache in Südtirol*. In: Egger, Kurt / Lanthaler, Franz (Hgg.): Die deutsche Sprache in Südtirol: Einheitssprache oder regionale Vielfalt. Wien – Bozen: Folio, 137-152.

Lemnitzer, Lothar / Zinsmeister, Heike (2006): *Korpuslinguistik*. Tübingen: Gunter Narr Verlag.

Masser, Achim (1982): *Italienisches Wortgut im Südtiroler Deutsch – droht eine Überfremdung?* In: Moser, Hans (Hg.): Zur Situation des Deutschen in Südtirol. Sprachwissenschaftliche Beiträge zu den Fragen von Sprachnorm und Sprachkontakt (= Innsbrucker Beiträge zur Kulturwissenschaft, Germanistische Reihe, Bd. 13). Innsbruck, Institut für Germanistik, 63-73.

Moser, Hans (1990): *Deutsche Standardsprache - Anspruch und Wirklichkeit*. In: IX. Internationale Deutschlehrertagung 1989. Tagungsbericht, Wien, 17-30.

Pernstich, Karin (1982): *Deutsch-italienische Interferenzen in der Südtiroler Presse*. In: Moser, Hans (Hg.): Zur Situation des Deutschen in Südtirol. Sprachwissenschaftliche Beiträge zu den Fragen von Sprachnorm und Sprachkontakt (= Innsbrucker Beiträge zur Kulturwissenschaft, Germanistische Reihe, Bd. 13). Innsbruck, Institut für Germanistik, 91-182.

Putzer, Oskar (1984): *Interferenz in Übersetzungen: Aspekte der Übersetzungsleistungen bei der Zweisprachigkeitsprüfung* (D.P.R. 752/76). Bozen: Assessorat für Schule und Kultur in italienischer Sprache.

Riedmann, Gerhard (1972): *Die Besonderheiten der deutschen Sprache in Südtirol* (= Duden Beiträge, Sonderreihe: Die Besonderheiten der deutschen Schriftsprache im Ausland 39). Mannheim-Wien-Zürich: Bibliographisches Institut.

Riehl, Claudia M. (2001): Schreiben, Text und Mehrsprachigkeit: Beitrag zu einer Teorie der mehrsprachigen Gesellschaft am Beispiel der deutschsprachigen Minderheiten in Südtirol und Ostbelgien. Tübingen: Stauffenburg Verlag.

Rizzo-Bauer, Hildegard (1962): *Die Besonderheiten der deutschen Schriftsprache in Österreich und Südtirol.* (= Duden Beiträge, Sonderreihe: Die Besonderheiten der deutschen Schriftsprache im Ausland 5). Mannheim-Wien-Zürich: Bibliographisches Institut.

Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees*. In Proceedings of International Conference on New Methods in Language Processing. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger> (Stand: Jänner 2011).

Sieber, Peter / Sitta, Horst (1994): *Sprachwandel - Sprachfähigkeiten*. In: Sieber, Peter (Hg.): Sprachfähigkeiten - Besser als ihr Ruf und nötiger denn je! (Reihe Sprachlandschaft 12), Aarau-Frankfurt a. M.-Salzburg: Sauerländer Aarau, 13-50.

Sitta, Horst (2001): *Sprachwandel, Sprachverfall - oder nur die ganz alltägliche Schlamperei?* In: Adamzik, Kirsten / Christen, Helen (Hg.): Sprachkontakt - Sprachvergleich - Sprachvariation. Festschrift für Gottfried Kolde zum 65. Geburtstag. Tübingen: Niemeyer, 345-356.

Spillner, Bernd (1992): *Deutsch-italienische Interferenzen bei Sprachkontakt und Mehrsprachigkeit.* In: Nelde, Peter H. (Hg.): It's easy to mingle when you are bilingual (= Plurilingua: Bilingualism and Contact Linguistics). Bonn: Dümmler, 173-186.

Tyroller, Hans (1986): *Trennung und Integration der Sprachgruppen in Südtirol*. In: Hinderling, Robert (Hg.): Europäische Sprachminderheiten im Vergleich. Deutsch und andere Sprachen (= Deutsche Sprache in Europa und Übersee 11.), Stuttgart.

## Internetadressen

Korpus Südtirol
<http://www.korpus-suedtirol.it> (Stand: Jänner 2011)
Einzelphänomensammlung zum Südtiroler Deutschen (Daum)
<http://www.kulturinstitut.org/fileadmin/user_upload/teaserbilder/veranstaltungen/Sprachstelle/SuedtirolerBesonderheiten_ListeDaum.pdf > (Stand: Jänner 2011)

C4-Initiative
<http://www.korpus-c4.org> (Stand: Jänner 2011)
Terminologieinformationssystem
*bistro* <http://www.eurac.edu/bistro> (Stand: Jänner 2011)
Online-Wörterbuch De Mauro – Paravia
<http://www.demauroparavia.it/26415> (Stand: März 2008; 2011: geänderte Zugriffs-modalitäten)

Vis-À-Vis
<http://www.korpus-suedtirol.it/korpus-suedtirol_II_de.htm> (Stand: Jänner 2011)
Digitales Wörterbuch der Deutschen Sprache
<http://www.dwds.de> (Stand: Jänner 2011);
<http://odo.dwds.de/wortprofil> (Stand: Jänner 2011)

Kookkurrenzdatenbank des IDS
<http://corpora.ids-mannheim.de/ccdb> (Stand: Jänner 2011)
Landesinstitut für Statistik ASTAT
<http://www.provinz.bz.it/astat/de> (Stand: Jänner 2011)

# Sprachdidaktik und Korpora

**Hans Drumbl – Freie Universität Bozen**

Abstract
Didaktische Erprobungen von Korpusinstrumenten in den drei Schulsprachen Südtirols, Deutsch, Englisch und Italienisch werden vor dem Hintergrund der weithin ungelösten Probleme des Zweit- und Fremdsprachererwerbs in gesteuerten Lernsituationen diskutiert. Ständige Bezugspunkte sind die in der internationalen Literatur kontrovers diskutierten Begriffe wie Muttersprachkompetenz, *native-like selection*, Idiomatizität, Fossilisierung in sprachlichen Bereichen von Phraseologismen, Kollokationen und Interjektionen. Fallstudien von Korpusrecherchen in den Arbeitsbereichen Editing und Sprachenlernen in unterschiedlichen Lernsituationen für Anfänger und für Fortgeschrittene werden in Bezug zu theoretischen Einblicken in den kognitiven Hintergrund von Sprachgebrauch und Spracherwerb diskutiert. Ausgehend vom *Competition Model* wird ein Erklärungsmodell für das bisher ungenügend untersuchte Phänomen der Stabilisierungsprozesse beim Spracherwerb zur Diskussion stellt.

## 1. Sprache als Erleben von Heimat und Fremde

„An der Spitze aller Kultur steht ein geistiges Wunder" – schreibt Jacob Burckhardt in seinen „Weltgeschichtlichen Betrachtungen" – „ein geistiges Wunder: die Sprachen, deren Ursprung, unabhängig vom Einzelvolke und seiner Einzelsprache, in der Seele liegt. (...) Eigentlich genügt zum Beweise schon das Erlernenkönnen fremder Sprachen überhaupt bis zum geistigen Gebrauch. So viele Sprachen, so viele *Herzen* besitzt man." (Vgl. die *tria corda* des Ennius)[1]

Wir können richtig nachvollziehen, wie das karge Zitat, das Ennius in indirekter Rede zu Wort kommen lässt – *Quintus Ennius tria corda habere sese dicebat, quod loqui Graece et Osce et Latine sciret* –, bei Burckhardt Früchte getragen hat. Und wer wollte Burckhardt widersprechen, dass Sprachen und Sprechen etwas mit dem Herzen der Menschen zu tun haben, mit den geistigen und emotionellen Kräften, die den Menschen als Gattung und als Individuum kennzeichnen? Wenn wir den Enthusiasmus in Burckhardts Worten zu den wissenschaftlichen Erkenntnissen zu Kognition und Emotion in Bezug setzen, zu neuen Erkenntnissen wie der Rolle der Amygdala, den vielfachen Verbin-

1 Jacob Burckhardt, Weltgeschichtliche Betrachtungen, 2. Kapitel: Von den drei Potenzen, 3. Die Kultur, Stuttgart, o. J., S. 55.

dungen, die vom Präfrontalen Cortex (Damasio 1994), dem Sitz der höheren kognitiven Fähigkeiten zu den motorischen Zentren und weit „hinunter“ bis in die ältesten Schichten des Gehirns führen, wo Unbewusstes und durch Emotionen Gesteuertes ihren Ort haben (Lieberman 1991, 2000), und wenn wir schließlich den Gedanken, dass wir die Sprache in unserem Inneren hegen, mit den Spiegelneuronen und der von ihnen ermöglichten Empathie weiterdenken[2], werden wir das Überschwängliche der Worte des großen Schweizer Kulturphilosophen gerne akzeptieren und keinen grundsätzlichen Gegensatz zu der uns heute leitenden wissenschaftlichen Sicht der Sprachverarbeitung darin sehen.

Ebenso kann es uns bei einem gewichtigen Wort Hugo von Hofmannsthals ergehen, eines Österreichers in einer schwierigen Zeit. Hofmannsthal war Zeuge eines Umbruchs und eines Verlustes ungeahnten Ausmaßes, aber er besaß auch eine besondere Gabe dafür, Auswege aus anscheinend ausweglosen Situationen zu finden. Der Verzicht auf die Dichtkunst des Lord Chandos war ein solcher Ausweg, das Schweigen des Karl Brühl, des „Schwierigen“, ein anderer. Gerade in solchen Zeiten blieb ihm jedoch unverrückt ein Bild der Sprache gegenwärtig, das noch heute auf uns wirken kann. 1927 berührt Hofmannsthal in seiner Rede an der Universität München, *Das Schrifttum als geistiger Raum der Nation* diesen Punkt :

> Nicht durch unser Wohnen auf dem Heimatboden, nicht durch unsere leibliche Berührung in Handel und Wandel, sondern durch ein geistiges Anhangen an allem sind wir zur Gemeinschaft verbunden. Hierdurch unterscheiden sich unsere alten europäischen Nationen von jenem jungen, nach außen mächtigen amerikanischen Staatswesen, in dem wir eine Nation in diesem Sinne noch nicht zu erkennen vermögen. In einer Sprache finden wir uns zueinander, die völlig etwas anderes ist als das bloße natürliche Verständigungsmittel; denn in ihr redet Vergangenes zu uns, Kräfte wirken auf uns ein und werden unmittelbar gewaltig, denen die politischen Einrichtungen weder Raum geben, noch Schranken zu setzen mächtig sind[3].

Die Worte Hofmannsthals haben das Gewicht eines Programms, wie wir es heute von der Spitze der EU gewohnt sind, zugleich aber auch geben sie eine kulturpolitische Begründung für die Achtsamkeit und die Pflege unserer Sprachen, der eigenen wie der neuen Sprachen, die in unser Leben treten. Keineswegs können wir bei Hofmannsthal eine Beschränkung auf das Eigene in einem eher eng gefassten regionalen Sinne erkennen, wie sie in Nadlers

2 Rizzolatti et al. 2009 gibt einen Überblick über die klinisch relevanten Studien. Siehe auch Ramachandran, V. S. / Oberman, L. M., “Broken Mirrors”, in: Scientific American, Nov. 2006.

3 Ich verweise auf das Faksimile des Erstdrucks http://www.ub.uni-bielefeld.de/diglib/2006/hofm_schrifttum/

Literaturgeschichte so markant in Erscheinung tritt. Es geht um die Funktion der Sprache für das Entstehen einer Gemeinschaft, um die Vertrautheit mit der „Kultur“, die in der Sprache ihren Ausdruck findet.
Ganz ähnlich äußert sich Isaiah Berlin, der große liberale Denker des vergangenen Jahrhunderts, wenn er in einer differenzierten Auseinandersetzung mit der Aufklärung und ihrer Tendenz, das Allgemeine anzustreben, einen kritischen Blick auf den Begriff einer Gemeinschaft ohne lokale Begrenzung wirft:

Es sei ein Fehler der Philosophen der Aufklärung gewesen anzunehmen, dass Männer und Frauen nach abstrakten Prinzipien und kosmopolitischen Wertvorstellungen und dem, was er einen „idealistischen, aber leeren doktrinären Internationalismus“ nannte, leben könnten.
Diese Ablehnung natürlicher Bindungen scheint mir edel, aber fehlgeleitet zu sein. Wenn die Menschen sich beklagen, einsam zu sein, meinen sie, dass niemand versteht, was sie sagen: Verstanden zu werden bedeutet, eine gemeinsame Vergangenheit zu haben, gemeinsame Gefühle, eine gemeinsame Sprache, gemeinsame Voraussetzungen, die Möglichkeit vertraulicher Mitteilung – kurz gesagt, gemeinsame Formen des Lebens zu teilen[4].

Das Teilen von gemeinsamen Formen des Lebens ist auch ein „Teilen“ im Bereich der Sprache. Ein Teilen, das als Hintergrund des Sprechens die Strategien der Sprecher lenkt, und das in der Sprache selbst als Form verankert ist. Helmuth Feilke hat diese „Form“ des Teilens in eindringlichen Analysen als „Common sense-Kompetenz“ herausgearbeitet und hat in diesem Zusammenhang auf einen ersten Ansatz zu dieser Erkenntnis bei Max Weber hingewiesen:

Das Bestehen einer ‘Sprachgemeinschaft’ bedeutet für uns nicht, daß massenbedingte Gleichartigkeit bei der Hervorbringung bestimmter Lautkomplexe existiere (das ist gar nicht erforderlich), auch nicht nur, daß der eine ‘nachahmt’, was andere tun, sondern vielmehr ein Verhalten bei ‚Äußerungen’, welches an bestimmten, innerhalb eines Menschenkreises durchschnittlich bestehenden Chancen, sich ‚verständlich’ zu machen, sinnhaft orientiert ist und daher diesen Effekt im Durchschnitt auch erwarten ‚darf’. (Weber zitiert nach Feilke, 1994, S. 78)

Die Menschen in einer Sprachgemeinschaft sind durch Erfahrungen miteinander verbunden, und dieser Erfahrungshintergrund ist es, der den Sprachgebrauch zum Erfolg führt. Die Erwartung des Erfolgs unterscheidet Sprecher im Umfeld einer Sprachgemeinschaft von Lernenden, die sich einer Sprache erst nähern. Erfolgreiche Lernende sind erfolgreich, wenn es ihnen gelingt, in genau diesem Sinn zu Teilnehmern der Sprachgemeinschaft zu werden.

4 Michael Ignatieff, Isaiah Berlin. Ein Leben, (1998), S. 374.

Man kann von einer Kompetenz sprechen, von einer Haltung, die das „Glücken" von Kommunikation erwartet, die mit dem Glücken rechnet. Vielleicht ist es gerade diese Erwartung, die das Gefühl vermittelt, „dazu zu gehören", als Sprecher einer (neuen) Sprache nicht mehr in der Rolle des Anfängers zu agieren, sondern als vollwertiger Teilnehmer einer Sprachgemeinschaft. Wohlgemerkt, ganz unabhängig von der Perfektion oder von der Fehlerhaftigkeit der einzelnen Äußerungen, die der Neuankömmling hervorbringt.
Die Teilnahme an der Sprachgemeinschaft kann also auch als Haltung beschrieben werden, als eine Erwartungshaltung plus einem bestimmten Grad von Sicherheit, dass die Erwartung auch erfüllt werden wird. Genau diese Haltung müssen Lernende der zweiten oder dritten Sprache erreichen. Ich sage „erreichen", korrigiere diesen Anspruch aber sogleich in Gedanken an ein denkwürdiges Wort von Valéry über die „Kunst des Erziehenden":

> Die Kunst des Erziehenden besteht darin, Aufmerksamkeit zu schaffen, sie freiwillig werden zu lassen, ihren Aufbau zu ermöglichen; sobald sie vorhanden ist, sie zu bewahren, ihre Funktionsweise zu kontrollieren und ihre Anwendung zu begrenzen. Man muß ihr schonend die *entsprechende* Nahrung geben – nie zu viel, nie zu wenig, ihr ein Ziel stecken und sicherstellen, daß es angestrebt wird. (Valery 5:607)

Und diese Haltung, mit dem „Glücken" des jeweiligen Sprachaktes zu rechnen, kann nicht gelernt sondern nur „erfahren" werden und das heißt, durch wiederholtes Üben gewonnen werden. Vielleicht muss man die Haltung als Lernender einfach voraussetzen und dann – unter der Voraussetzung, dass das Vorhaben gelingen wird – immer wieder übend anwenden. Systematisches Arbeiten mit Korpusinstrumenten ist ein Mittel – so möchte ich unsere Erfahrungen zusammenfassen –, genau diese Haltung zu erfahren und zu erproben und in immer neuen Kontexten zu üben[5].

## 2. Schreiben im beruflichen Umfeld

Das Erlernen der Muttersprache bis zu dem Niveau, das durchschnittliche Sprecher erreichen, ist eine der großen kulturellen Leistungen des Menschen. Das zu erreichen, was jedes normal aufwachsende Kind in seiner Muttersprache anscheinend mühelos erreicht, aber auf andere Weise als die natürliche des Erstspracherwerbs, nämlich durch Lernen als Jugendlicher oder als Erwachsener, scheint hingegen ein unerreichbares Ziel zu sein.

5 Als Grußwort und Einleitung zur Tagung vorgetragen. Im folgenden Teil werden didaktische Erfahrungen mit Korpusinstrumenten referiert, die zu unserer gängigen Praxis gehören. (Drumbl 2002, 2005, 2009, Zanin 2007).

Wenn man aber das anscheinend unerreichbare Ziel der sehr guten Sprachkompetenz näher zu bestimmen versucht, verändert sich die Perspektive, denn was anderes kann das Ziel des Erlernens einer fremden Sprache sein, als diese so zu sprechen, wie Menschen sie sprechen, für die sie nicht eine fremde, sondern die eigene Sprache ist? Und diese Sprache, dieses Sprechen der Muttersprachler ist von einer immensen Vielfalt von Varietäten, Varianten und qualitativen Abstufungen gekennzeichnet. Ist das, was diese Menschen beim Gebrauch ihrer Sprache meist ohne Anspruch auf Perfektion vollbringen, so unvorstellbar weit entfernt von dem, was man als erwachsener Lernender lernen kann, lernen könnte, wenn das Erlernen der Sprache glücken würde?
Man muss das lernen, was die anderen können. Dazu gehört, dass man auf das achtet, was die anderen *tun*, die Menschen, die in zweisprachigen Gebieten nebeneinander leben, die einer für den anderen in jedem Moment Vorbilder sein können/könnten für das Sprechen und das sprachliche Handeln.
Vielleicht liegt ein grundlegender Fehler in der Haltung, überhaupt nicht hinzusehen, nicht hinzuhören, wenn andere Sprecher die Sprache gebrauchen. Lernende der Sprache handeln oft als Wartende, sie, die Lernenden, seien ja keineswegs schon bereit zu handeln, sie müssten erst „lernen". Im beruflichen Umfeld ist der Verzicht auf die implizite Hilfe durch das Beobachten von bereits erfolgten Sprechakten kompetenter Sprecher besonders auffällig – und schädlich. Der hier vorgelegte Einzelfall soll als Argument dafür betrachtet werden, die Alternative zu erproben, das unerreichbar scheinende Ziel der idiomatischen Kompetenz (*native-like selection*) so zu behandeln als sei es ein ganz gewöhnliches Ziel, das ohne weiteres angestrebt werden kann und das keineswegs die unüberwindliche Trennlinie darstellt zwischen Muttersprachlern und Nicht-Muttersprachlern.[6]
Man muss das lernen, was die anderen können: das ursprüngliche Prinzip des Lernens bei einem „Meister", die Störz der Lehrlinge. Das ist keine Ausbildung für Schüler, die „für später" lernen, sondern für Lehrlinge, die in der Werkstatt bereits arbeiten und lernen.
Der Weg kann durch eine Fallstudie exemplifiziert werden, die einen Moment des Misslingens betrachtet. Dazu wähle ich einen Moment der täglichen Arbeit eines Teilnehmers unserer Sprachgemeinschaft in Südtirol und zwar im engeren Rahmen der Freien Universität Bozen, der dreisprachigen Universität mit ihren besonderen Aufgaben im Bereich der Mehrsprachigkeit. Ein kleiner Moment im täglichen Ablauf, statistische Daten zur Universität müssen mit Überschriften und einfachen Begleittexten versehen werden. Diese Kurztexte werden in den laufenden Text eingebunden, das Dokument erreicht die ver-

6 Native speaker vs. Non native speaker (NS vs. NNS) in der internationalen Diskussion.

schiedenen Gremien, wird dort diskutiert, schließlich approbiert, gedruckt und verteilt. Im Normalfall wird dem sprachlichen Ausdruck in keiner der auf den Akt des Schreibens nachfolgenden Etappen große Aufmerksamkeit geschenkt. Daher ist der erste Moment, der Moment der Verschriftlichung so wichtig. Daher müssen die Mitarbeiter der Institutionen in der mehrsprachigen Arbeitsumgebung mit besonderer Sorgfalt mit der Sprache umgehen – und sie müssen darin ausgebildet werden. Das ist der Angelpunkt, der diese Episode mit dem Thema unserer Tagung verbindet.
Der kleine Moment des Schreibens betrifft die Überschrift zur Statistik der Herkunft der Studenten der Freien Universität Bozen. Über dem gewohnten Tortenbild der Daten kann als Überschrift kaum mehr und etwas anderes stehen als: Herkunft der Studierenden. Die gender-neutrale Form hat sich in den letzten Jahrzehnten gegenüber dem traditionellen Sammelbegriff Studenten durchgesetzt. In unserem Internetkorpus spiegelt sich diese Entwicklung in der Zahl der Belege wieder: 43 „Studierenden" stehen 13 „Studenten" gegenüber. Im Alltag der Kommunikation sehen die Zahlen allerdings etwas anders aus (Google-Recherche vom 27. Juli 2010): 20.800 (Herkunft der Studierenden) gegen 18.500 (Herkunft der Studenten).
Das gibt einen ersten Anstoß zum Auswählen, wobei in diesem Fall die Wahl kaum als eigenständiger individueller Akt erfahren wird, sondern als Teilhabe einer von der Universitätsgemeinschaft vertretenen Haltung gleichsam vorgeschrieben ist. Diese Haltung zeigt sich darin, dass in ähnlichen Texten der Verwaltung und der Öffentlichkeitsarbeit gender-neutrale Ausdrücke bevorzugt werden.
Die Entscheidung zwischen zwei parallel überlieferten Varianten wäre kaum der Rede wert, gäbe es in diesem Text nicht eine überraschende Abweichung vom universell dokumentierten Sprachgebrauch. Im Text unserer Universität steht nämlich etwas anderes. Dort steht Herkunft der gesamten Studierenden. Eine wohl durch Nachdenken gefundene „hyperkorrekte" Präzisierung, die nicht nur dem Sprachgebrauch widerspricht, sondern auch der sprachlichen „Logik" der Aussage selbst, denn die Antwort auf die Frage nach der Herkunft der gesamten Studierenden könnte nur lauten: „aus 15 Ländern", „von drei Kontinenten" oder „vom Planeten Erde".

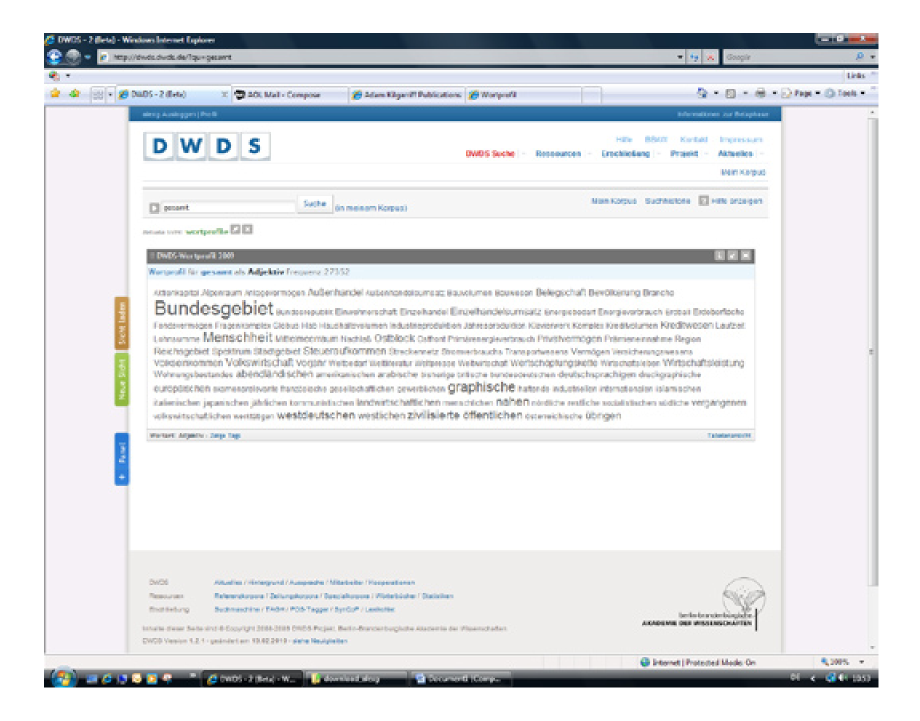
Von „den gesamten Studenten" ist nur im Zusammenhang mit Untergruppen der Studierenden die Rede (Belege dewac)[7]:

7 Die Belege sind nicht redaktionell redigiert. Die Satzzeichen sind durch Leerzeichen vom Text getrennt. Das ist bei Korpusanfragen im Phrase-Fenster zu beachten.

Die Zahl der **gesamten** Studenten nach Studiengang:
Der Prozentanteil gibt an , wie groß der Anteil der Studenten dieses einen Studiengangs gemessen an den **gesamten** Studenten innerhalb der Lehreinheit ist.
Diese vier Sprecher bilden den Sprecherrat , dem die Vertretung der Interessen der gesamten **Studierenden** auf Universitätsebene obliegt.
An den Elitehochschulen ( „Grandes Ecoles“ ), die gemeinsam knapp fünf Prozent der gesamten **Studierenden** ausbilden , wird durch Aufnahmeprüfungen gefiltert.
Dieses Feld enthält die Menge der gesamten **Studierenden** für die ausgewählte Studienrichtung.

Der „innovative“ Ausdruck im Bozener Uni-Dokument wird von Google nur in zwei Belegen nachgewiesen, nämlich genau in zwei Texten unserer Universität und nirgendwo sonst auf der Welt. 20 Sekunden Arbeitszeit hätten ausgereicht, um den Zweifel zu beheben – falls der Zweifel aufgetaucht wäre. Die Stelle ist aber weder beim Schreiben noch beim Lesen aufgefallen. Die kleine Gemeinschaft im sprachlichen Biotop der mehrsprachigen Universität Südtirols hat sich mit solchen Ausnahmen zum gängigen Sprachgebrauch abgefunden und hat ein weiteres Fallbeispiel zu den bereits bestehenden hinzugefügt. Eine Enklave eines abweichenden Sprachgebrauchs wird geschaffen und für weitere Nutzung geöffnet.
Die Menschen in dieser Enklave handeln anders als die „normalen“ Mitglieder der Sprachgemeinschaft, sie haben keinen Anteil am Bewusstsein der sprachlichen Muster, denen das alltägliche Sprechen seine Routinen verdankt. Vorgefertigte Muster, die von den Teilnehmern der Sprachgemeinschaft geteilt werden, erleichtern die Anstrengung des Schreibens und des Sprechens ungemein. Ohne diese Fertigteile müsste jeder Gedanke auf eine individuell gestaltete Art und Weise ausgedrückt werden – ein kognitiver Aufwand, dem die Menschen im Normalfall ausweichen, um ihre Ressourcen dem Aufbau des Gemeinten, der Strukturierung ihres Redebeitrags zu widmen, und der Aufmerksamkeit auf die Sprechakte des Gegenübers.
Das Adjektiv „gesamte“ ist fehl am Platz und dieses Fehl-am-Platz-Sein betrifft das Adjektiv selbst. In diesem Fall sind die Gebrauchsbedingungen – die *constraints* – überzeugend nachzuweisen, es sind spezifische Bereiche im Umfeld politischen und gesellschaftlichen Überblicks, der Verwaltung, bevorzugt in fachsprachlichen Bereichen, in denen das Adjektiv mit hoher Frequenz gebraucht wird.
Hier nach den Wortprofilen des dwds abgebildet[8]:

8 Siehe dazu Geyken in diesem Band.




Der Verfasser der Broschüre wählt das Adjektiv „gesamt“ ohne an die traditionellen Gebrauchsbeschränkungen und Muster zu denken. Es sind dies die Muster, auf die er sich beim normalen Sprechen stützen würde und auch tatsächlich stützt. Hätte der Verfasser eines Textes Anlass, Wörter wie „Bundesgebiet“, „Außenhandel“ oder „Belegschaft“ zu verwenden, dann läge auch der Schritt zur Verwendung von „gesamt“ nahe. In Kontexten, in denen keine im Wortprofil nachgewiesene „Begleiter“ des Adjektivs „gesamt“ vorkommen, sinkt die Wahrscheinlichkeit entsprechend. Wer den Bericht über die Freie Universität Bozen verfasst oder übersetzt hat, hantiert an dieser Stelle ganz offensichtlich mit einem Wort in Isolation, wie Wörter beim Übersetzen der rudimentären Form des Wort-zu-Wort-Übersetzens vorkommen. Das Wort hat alle latenten Bezüge zu Kontexten, zu Kollokationen und zu Gebrauchsformen und Mustern verloren.

Der zufällig gefundene „Fehler“ ist als Indiz für ein Verhalten zu deuten, das im Umfeld der Mehrsprachigkeit tagtäglich vorkommt, das Isolieren von Wörtern, das in der Folge zu einer Überangestrengtheit beim Sprechen und beim Schreiben führt, und zu einer Art Hyperkorrektismus in eigener Verantwortung. Dass dabei eine ungewöhnliche und ungewohnte geistige Anstrengung

erbracht werden muss, ist das eigentliche Problem, nicht der Fehler als solcher, der in Folge dieser fehlgeleiteten Anstrengung entstanden ist. Hyperkorrektismus und Verunsicherung gehen Hand in Hand. Wer so genau kontrollieren will, ist sich der Gefahr des fehlerhaften Sprachgebrauchs bewusst. Er/sie ergänzt die normale Aussage durch das Adjektiv „gesamten“ aus Angst vor einem ungenügenden Ausdruck, im Bewusstsein, „vollständig“ denken und schreiben zu müssen, als ob die Last der Sprache allein auf den Schultern des Einzelnen läge, der sich gerade voller Skrupel an das Verfassen einer kleinen Überschrift macht. Warum wird ein Moment der Unsicherheit nicht zu einem Moment des Lernens, zu einem Blick auf die *best practice*, auf Vorbilder, die ohne großen Aufwand aufgerufen werden können?
Der Kontakt mit den Sprachen am Arbeitsplatz erscheint hier als Handeln in der Isolation der eigenen, abgeschotteten Sprache, ein Handeln, als ob es keine anderen Menschen gäbe, die unsere Sprache verwenden und keine Texte, in denen dieses sprachliche Handeln aufgezeichnet wäre.

## 3. Stabilisierungsprozesse

Mit der Angst vor dem Fehler ist ein zentraler Punkt des Spracherwerbs und des Sprachunterrichts in der Schule angesprochen. Zur Angst kommt die täglich erneuerte Erkenntnis des Misslingens von Mitschülern und von Menschen im sozialen Umfeld, denen es nicht gelungen ist – und nicht gelingen will – die fremde Sprache zu lernen. Was den anderen neben mir nicht gelingt, warum sollte es „mir“, dem Einzelnen gelingen, der sich gerade daran macht, die Sprache zu lernen?
Sprachenlernen in der Schule bringt in der Tat – vor allem für schwächere Schüler – neben den entmutigenden Beispielen weitere Nachteile mit sich. Der Syllabus ist nach den Bedürfnissen der Institutionen konzipiert, die bestimmte Ergebnisse anstreben und durchsetzen müssen. Die einzelnen Lernschritte setzen Vorkenntnisse und *skills* voraus, die nicht von allen Schülern zum Zeitpunkt der Behandlung des Programmpunktes in ausreichendem Maße beherrscht werden. Dadurch entstehen immer neue Differenzen in der Sprachkompetenz, aber auch in der Bereitschaft, die neue Sprache zu lernen (Han 2004, 2009). Differenzen, die immer wieder dieselben Schüler ausgrenzen.
Dieses „Zurückbleiben“ der einen, das „Nicht-Mitkommen“ mit den anderen Lernenden, denen das Lernen gerade im richtigen Moment gelingt, ist eine Grunderfahrung der Lernenden wie der Lehrenden von Fremdsprachen in der Schule.

Die andere Grunderfahrung der meisten Lernenden ist die Erkenntnis, dass es ihnen nicht gelingt und wohl auch in Zukunft wahrscheinlich nicht gelingen wird, alle grammatikalischen Feinheiten zu erkennen, zu verstehen und zu gebrauchen, so wie es ihnen nicht gelingt, die Feinheiten der Aussprache, vor allem der Intonation wahrzunehmen und selbst adäquat auszudrücken. Immer wieder machen sie dazu noch die Erfahrung, dass ihr Wortschatz unvollständig ist. Selbst nach jahrzehntelangem Aufenthalt im Land und aktiver Teilnahme an den Aktivitäten der Menschen dieser Sprachgemeinschaft erkennen sie immer wieder Lücken in ihrer Sprachkompetenz, so dass es den wenigsten gegeben ist, die zweite oder dritte Sprache im Gefühl zu gebrauchen, von jeder störenden Unzulänglichkeit frei zu sein.

Je fortgeschrittener die Sprachkompetenz und die Sprachbewusstheit des Sprechenden, desto klarer die Erkenntnis, dass seine Sprachkompetenz auf eine ganz eigene, auf eine persönliche Weise „lückenhaft“ geblieben ist: Manchmal fehlen ganze Bereiche des Alltags, manchmal fehlen bestimmte Register, manchmal geht eine deutliche Trennlinie zwischen mündlicher und schriftlicher Kompetenz, und andere Fälle mehr.

Dieser Zustand ist stabil. Und in den meisten Fällen bleibt er stabil, ein Leben lang. Der einzelne Sprecher sieht die eigene Sprachkompetenz im Normalfall aber nicht im Vergleich zur Kompetenz der Anderen. Er misst sich nicht an den anderen und hat kein konstantes Gefühl des Unbehagens an seinen beschränkt(er)en Möglichkeiten des Ausdrucks. Die Beschränkungen werden aber von den verunsicherten Lernenden wahrgenommen und als Indizien eines Defizits empfunden.

Fossilisierung wurde in der Forschung oft als Symptom des verhinderten Erwerbs einer Sprache, des verhinderten Weitergehens beschrieben: „that learners typically stop short of target-like attainment“ (Han 2006:137 als Gemeingut der Forschung zitiert), so dass das Beharren auf einer bestimmten Stufe der Interimssprache auch generell als „incompleteness“ der Sprachkompetenz der Lernenden bezeichnet wird (Schachter 1988, 1996). In einer lernerzentrierten Sicht, die sich nicht auf didaktische Strategien beschränkt, sondern die auch die Haltung des sprechenden Menschen zur Sprache und seinen Umgang mit Sprachen im Alltag beobachtet, erscheint die Fossilisierung als Phänomen, das jeden Betroffenen auf eine ganz eigene, idiosynkratische Weise trifft. Bestimmte Merkmale können bei vielen Sprechern vorkommen, einige nur bei wenigen, die genaue Verteilung der Merkmale aber ist unbestimmbar. Sie variiert auch innerhalb der sprachlichen Aktivitäten des Einzelnen, ist von Faktoren abhängig wie der Modalität, der Registerebene, des Grades an Aufregung, an Vorbereitung, an Sachkenntnis, und all der Faktoren, die bei kommunikativen Anlässen eine Rolle spielen können.

In vergleichender Sicht erscheinen die „Mängel" von Fall zu Fall als Eigenheiten eines Sprechers, der mit Stress weniger gut umgehen kann, der mit der hohen Stilebene weniger gut vertraut ist, der im Zusammenfassen komplexer Texte wenig Übung hat, und dessen Stimme ungeübt ist im öffentlichen Vortrag. Es ist eine Fülle von Einzelheiten des Sprechens, des Vortragens, des Komponierens, des Meinens und des Ausdrückens von Gemeintem, die dem aufmerksamen Sprecher aus vielen Momenten des Alltags vertraut ist.

## 4. Das Konzept „Resonanz" und seine neuronale Grundlage

Beim Zusammenleben zweier oder mehrerer Sprachen werden durch wiederholten Sprachgebrauch die Einzelteile mit anderen Sprachteilen vernetzt. Wörter und Wortgruppen verbinden sich in einem Netz von Wahlverwandtschaften zu Einheiten, die zusammen mit anderen ein eng geflochtenes Gemeinsames an Sprache bereitstellen, das nachhaltig im Gehirn gespeichert ist. Bei jedem Aufrufen eines Einzelnen werden zugleich die mit-verbundenen sprachlichen Elemente aktiviert, die als „Resonanz" zu den jeweils aufgerufenen sprachlichen Elementen dazukommen[9].
Mit Hilfe der Resonanz wird ein Raum für jede neue Sprache geschaffen, der im Folgenden von anderen Sprachräumen gegen ungewollte Einflüsse (Interferenzen) gleichsam abgeschirmt wird. Je mehr solcher miteinander verbundener Wörter bereits im Gehirn gespeichert sind, desto leichter finden Neuankömmlinge den Ort für sich selbst, desto schneller kann man sie sich merken.
Beim simultanen Spracherwerb von Kindern, aber auch beim Spracherwerb im fortgeschrittenen Alter, wird jede Sprache im Gehirn „verwurzelt", und im Fall von Mehrsprachigkeit bilden sich getrennte sprachliche Systeme aus, die sich im Lauf des Spracherwerbs immer stärker voneinander abgrenzen. Der Übergang von der einen zur anderen Sprache wird dabei von eigenen neuronalen Strukturen geregelt, die von Forschern des „San Raffaele" in Mailand genau im Gehirn lokalisiert werden konnten (Abutalebi et alii, 2007; vgl. Drumbl 2009: 76-77). Die Erforschung der „Switches" beim mündlichen Sprachgebrauch hat ein Ergebnis erbracht, das von den Autoren der Studie sogleich in seiner Tragweite für den Unterricht der Zweitsprache erkannt wurde. Entgegen den Erwartungen war der Aufwand an kognitiver Kontrolle beim Wechsel von der Zweitsprache zur Erstsprache höher als umgekehrt. Es handelte sich dabei um Versuchspersonen, die in jahrelangem Kontakt mit der Zweitsprache leben und die ihre Erstsprache weniger benutzen als die Zweit-

9 Bates/Goodmann 1999, McWhinney 1987. In diesem Zusammenhang ist auch die Darstellung der Priming-Effekte von Bedeutung; vgl. Hoey 2005.

sprache. Damit ist erwiesen, dass neuronale Strukturen durch Kontakt mit der Sprache geschaffen werden, mit anderen Worten, dass die Sprachen und ihre neuronalen Grundlagen durch Lernen entstehen. Das Gehirn wächst beim Erlernen der neuen Sprache mit. Auch im Erwachsenenalter.
Fremdspracherwerb verläuft oft unter extrem ungünstigen Ausgangsbedingungen. Wenn die zweite Sprache zu schwach ist, um den Prozess der Resonanzbildung auszulösen, besteht die Gefahr, dass die Voraussetzungen für diesen Prozess überhaupt nicht geschaffen werden, so dass die Lernenden eigentlich immer am Anfang stehen und bei Null beginnen. Daher ihr Bedürfnis, sich mit Hilfe der Wort-für-Wort-Übersetzung ein Minimum an Sicherheit zu verschaffen, das die nicht erreichbare Sicherheit durch die systematische Ausbildung der Sprachkompetenz kompensieren soll. In einem Artikel, der auf jahrelangen Vorarbeiten von Elizabeth Bates aufbaut, ist diese neue Sicht der Sprache und der Sprachkompetenz so beschrieben:

When the child's two languages are roughly similar in dominance or strength, each system generates enough system-internal resonance to block excessive transfer. However, if one of the languages is markedly weaker, then it will not have enough internal resonance to block occasional transfer. The situation is very different for L2 learners, since the balance between the languages is then tipped so extremely in favor of L1. In order to permit the growth of resonance in L2, learners must apply additional learning strategies that would not have been needed for children. These strategies focus primarily on optimization of input, promotion of L2 resonance, and avoidance of processes that destroy input chunks. (Hernandez et alii 2005: 18).

Die Resonanz entsteht durch das Vernetzen von Wörtern. Wortschatz und Grammatik spielen dabei zusammen eine Rolle und können daher auf keiner Ebene des Sprachgebrauchs und der Sprachbetrachtung voneinander getrennt werden, weder im theoretischen Modell der Sprache, noch im Hinblick auf die Prozesse des frühkindlichen Spracherwerbs, noch als Grundlage des Erlernens von Sprachen in fortgeschrittenerem Alter. Grundelemente der Sprache sind „Konstruktionen“ (Goldberg 2003, 2006, Feilke 2007), typischerweise Mehr-Wort-Ausdrücke, die schon beim alltäglichen Gebrauch eine zusammenhängende Einheit bilden. Beim Gebrauch von Konstruktionen werden sowohl die einzelnen Elemente als auch die idiomatisch geprägten Ausdrücke gelernt und beim Sprechen „wiederholt“. Es kommt zu einer Fülle von Begegnungen mit der neuen Sprache. Sprache wird als statistisch relevantes Phänomen erfahrbar, und die Muster, in denen sprachliche Konstruktionen im Alltag des Sprachgebrauchs in Erscheinung treten, werden als relevant erkannt und erleichtern das Erlernen der Grundelemente, aber auch das Fortschreiten im Umgang mit

sprachlicher Komplexität in einem Lernprozess, der sich über die gesamte Lebensspanne der Menschen erstreckt.

## 5. Förderung der „Resonanz" durch Lernen mit Korpusinstrumenten

Das Wachstum der Resonanz muss im Unterricht bevorzugt gefördert werden. Dazu gehört, dass das Lernen auf der Grundlage der Erstsprache unterbunden und das Lernen von fertigen Einheiten, *chunks*, unterstützt und gefördert werden. In dieser Perspektive kann die Arbeit mit Korpusinstrumenten hervorragende Dienste leisten. Die Suche in Korpora, selbst-motiviert und als entdeckendes Lernen gestaltet, wird zu einem wirkungsvollen Instrument des Lernens. Gelernt werden sollen Mehr-Wort-Verbindungen, und zwar nicht nur die phraseologisch unmittelbar in ihrer funktionalen Evidenz erfahrbaren pragmatisch verwertbaren Einheiten, sondern auch andere, mehr oder weniger stabile Verknüpfungen, die in Texten tatsächlich aufscheinen und den – oft noch unerkannten – Kern der Sprachkompetenz des Einzelnen darstellen und zugleich als Bindeglied zwischen den Sprachteilnehmern fungieren (Wray 2001, Belica 2001, Heringer 1999, Bubenhofer 2009).
Geschult wird bei diesem abgeleiteten sprachlichen Handeln des „Suchens" nach Korpusbelegen, die Aufmerksamkeit nicht nur auf die Form, auf wiederkehrende Muster zu lenken, sondern auch ein Gefühl dafür zu entwickeln, mit welcher „Wahrscheinlichkeit" bestimmte Zusammenstellungen tatsächlich aufscheinen. Es handelt sich gleichsam um eine Metaebene der Resonanz-Phänomene, die unbewusst vor sich gehen.
Je stärker beim Gebrauch einer Zweitsprache die Resonanz ausgebildet ist, desto leichter können andere Wörter, „Neuankömmlinge", sich in das bestehende Netz einfügen und ihrerseits wieder, als neu gespeicherte Wörter, andere, neu hinzukommende Wörter verankern. Je weniger Wörter gespeichert sind, desto schwerer ist es Wörter zu lernen.
Daraus ergibt sich eine Definition des kompetenten Sprechers, bzw. der muttersprachlichen Kompetenz: Das Maß der Kompetenz ist der Grad an Verknüpfung der sprachlichen Elemente im Resonanzraum der betreffenden Sprache. Erfolgreich ist der Erwerb der Zweitsprache dann, wenn der zweiten Sprache in der Konkurrenz zur Erstsprache genügend Aktionsraum zur Verfügung steht, um die „Resonanzen" aufzubauen, die als funktionaler Hintergrund für den Sprachgebrauch dienen.
Das Modell der Resonanz als Grundlage des Sprachgebrauchs und des Spracherwerbs gibt dem Begriff der Fossilisierung eine neue Grundlage. Fossilisierung ist

anders als „native-like selection". Fossilisiertem Sprachgebrauch fehlt nicht eine bestimmte Menge an Regeln oder an sprachlichen Mitteln, der Sprecher geht mit den verfügbaren sprachlichen Mitteln anders um. Die normale „fluency", die „native-like selection" ist kein Zustand, sondern eine *Bewegung* im Feld der Resonanz, eine Bewegung mit immer neuen Veränderungen, Verstärkungen bestimmter Resonanzen, Abschwächen anderer, die dann beim Sprechen oder beim Schreiben zu den jeweils aktuellen Priming-Effekten führen.

Priming und Resonanz sind die zwei Pole, zwischen denen die native-like selection angesiedelt ist, wo sie sich entfaltet und wo sie sich zu erkennen gibt. Das Ergebnis solchen Sprechens, eines Sprechens, das als normales Ergebnis „normal" gute Äußerungen hervorbringt? Diese Sprachkompetenz ist „gut", allerdings auf eine unvorhersehbare Weise gut. Der einzelne Sprechakt, die einzelne Äußerung haben keine genormte Qualität, die „erreicht" werden könnte, die Qualität zeigt sich beim Aktualisieren des Sprechaktes und sie ist unvorhersehbar.

Die Leichtigkeit, mit der diese Qualität erreicht wird, gehört zu ihr wie das Staunen des Zuhörers oder Lesers, der sie wahrnimmt.

Diese konsolidierte Erfahrung, durch Selbstbeobachtung und jahrzehntelange Beobachtung von Lernenden gewonnen, erhält neue Konturen durch Beobachtungen aus dem Bereich des Lernens mit Korpora.

Dem Bild der „unvorhersehbaren hohen Qualität" beim Sprachgebrauch entspricht die zufällig gewonnene Einsicht beim Lesen von besonders auffälligen, „brauchbaren" Korpusbelegen. Diese glückhaften Funde haben mit dem *Serendipity*-Prinzip zu tun, das von Grund auf zur Suche in Korpusbelegen gehört.

Das im *Serendipity*-Prinzip enthaltene Finden entspricht wiederum dem Entdecken oder dem Bestimmen von Prototypen im Umgang mit heterogenen Daten. Das diesen Momenten innewohnende Ordnungsprinzip ist an verschiedenen Stellen greifbar. Immer geht es um ein Zentrum, das in loser Weise mit Daten verbunden ist, die mit diesem Zentrum einen Bezug aufbauen, oder durch einen Bezug verbunden sind.

Grundprinzip ist die Vernetzung, die Verknüpfung von Daten.

Wie können wir uns vor diesem Bild das Gegenbild des Sprechens im Rahmen einer fossilisierten Sprachkompetenz vorstellen? Beim fossilisierten Sprechen gibt es nichts Unvorhersehbares, also auch kein „Glücken" in besonderen Momenten, keine Bewegung auf dem Feld der Resonanz mit immer neuen Veränderungen. Die Beziehungen sind festgefügt und unveränderlich. Das entscheidende Merkmal scheint dabei nicht einmal das zu niedrige Niveau zu sein, auf dem sich diese Erstarrung eingependelt hat, sondern die Tatsache,

dass überhaupt eine Erstarrung eingetreten ist, die den Platz der Bewegung eingenommen hat. Es gibt fossilisiertes Sprechen auf sehr hohem sprachlichem Niveau. Gleichsam ein fehlerfreies fossilisiertes Sprechen.

## 6. Sprachen lernen mit Korpusbelegen

Warum Lernen mit Korpora? Und die Antwort: Um eine Haltung einzuüben. Eine Haltung, die zumindest beim Zweitsprachunterricht unverzichtbar ist. Eine Haltung, die auch beim Erlernen einer Fremdsprache in der Schule sinnvoll und fruchtbar ist, da sie autonomes Lernen ermöglicht und fördert. Eine Haltung, die Lernen zu einem als sinnvoll erfahrbaren Tun werden lässt.
Geschult wird beim sprachlichen Handeln des „Suchens“, die Aufmerksamkeit nicht nur auf die Form, auf wiederkehrende Muster zu lenken, sondern auch ein Gefühl dafür zu entwickeln, mit welcher „Wahrscheinlichkeit“ bestimmte Zusammenstellungen tatsächlich aufscheinen.
Arbeitsinstrument sind Korpora deutscher Gegenwartssprache, darunter auch sehr kleine Korpora (Ghadessy / Henry / Roseberry, 1996), die mit einer auf das Wesentliche beschränkten Suchsoftware befragt werden.

SketchEngine Korpus Selbs Justiz: Lemma bleiben

| | | |
|---|---|---|
| Firner Direktor geworden , aber für mich | **blieb** | er Kortens alerter Assistent . „ |
| brachte der Kellner einen Rohkostsalat . „ Ich | **bleibe** | bei meinem Apollinaris . Für dich habe |
| meinem Kater , erzählen können und ließ es | **bleiben** | . Ich trank den Mokka aus , und Korten |
| förmlich werden . Thomas wäre hartnäckig | **geblieben** | . Aber Danckelmann unterbrach : „ Es hat |
| mehr auffällig . Aber nach meiner Meinung | **bleibt** | der eine Bombe , die weitertickt . „ Das |
| sprang auf . „ Kann leider nicht länger | **bleiben** | . Bin auf ein Uhr zum Personenschutz „ |
| auf dem Werksgelände . Auf dem sollte ich | **bleiben** | ? Ich wollte mit Firner reden . In seinem |
| Betriebe im Rahmen des an sich Zulässigen | **bleiben** | , aber das Wetter damit nicht mehr fertig |
| vom Büfett zu seinem Tisch steuerte . Er | **blieb** | kurz stehen . „ Grüß Sie , haben Sie etwasZur |
| auf einem Laborschemel Platz zu nehmen und | **blieb** | selbst mit hängenden Schultern vor dem |
| nicht oft , daß ich vormittags zu Hause | **bleibe** | . Nicht nur , weil ich viel unterwegs bin |
| Western dreht und zeigt - ich bin ihnen treu | **geblieben** | . Es war zehn Uhr . Ich hatte , Heaven‘s |
| ich will mich bemühen , verständlich zu | **bleiben** | . Wir haben dafür gesorgt , daß auf eine |
| wäre froh , wenn Sie mit uns in Kontakt | **blieben** | . Setzen wir uns Montag vormittag wieder |
| gut verplombt , daß das nicht unbemerkt | **bliebe** | . Die andere Einbruchstelle ist die Sammelstelle |
| die kapitalistische Industrie sein . Dann | **bleibt** | für uns vom Staat die Rolle der Russen |
| zum Baden gehen würde , müßte ich draußen | **bleiben** | . Schade , ich hatte mich darauf gefreut |
| „ das muß aber bitte unter uns | **bleiben** | „ , einen Ausdruck des aktuellen Stands |

| | | |
|---|---|---|
| ich deinen Fall gelöst habe , aber es auch | **bleiben** | lassen . „ Nicht so empfindlich , mein |
| eine schöne blonde Nazisse war und auch | **blieb** | , bis sie zur vollschlanken |

Schon der erste Beleg erweist sich als äußerst produktiv: „... Firner Direktor geworden, aber für mich blieb er Kortens alerter Assistent“. Die polaren Rollen „Direktor“ und „Assistent“ sind als Internationalismen spontan verständlich, dadurch kann die semantische Funktion des „bleiben“ inferenziell mit großer Sicherheit erschlossen werden. Der zweite Beleg, „Ich bleibe bei meinem Apollinaris. Für dich habe....“ weckt das Interesse an der Fortsetzung:

Es waren zwei Gedecke aufgelegt , und als wir uns setzten , brachte der Kellner einen Rohkostsalat . „ Ich **bleibe** bei meinem Apollinaris . Für dich habe ich einen Chateau de Sannes bestellt , du magst ihn doch . Und nach dem Salat einen Tafelspitz ? „ Mein Lieblingsgericht . Wie nett von Korten , daran zu denken .

Ein klar strukturiertes Frame mit ausgewählten sprachlichen Elementen, die zusammen passen. Beim Durchforsten der KWIC-Liste springt das Auge von Beleg zu Beleg, sucht nach Interessantem, Auffälligem und setzt sogleich die unterschiedlichen Beispiele zueinander in Beziehung.
Die Suche ist funktional im Hinblick auf das Erlernen der Sprache. Was bei der Suche „funktionieren“ soll, sind ja nicht die von der Software gesteuerten Suchprozesse, sondern die Lernprozesse im Hirn des Lernenden. Mit Hilfe der Suche sollen Lernprozesse aktiviert werden, die ohne die Hilfe des externen Mittels nicht (so leicht / so systematisch) aktiviert werden könnten. Die Lernenden sollen *lernen*, sich der Sprache auf eine Art und Weise zu nähern, die es ihnen ermöglicht – die sie dazu führt – immer wiederkehrende Muster wahrzunehmen, darunter auch die Muster der formelhaft verbundenen Wörter, der *Zusammenstellungen*, die mit größerer statistischer Wahrscheinlichkeit auftreten als andere, die ebenso gut möglich wären.
Die Lernenden können dabei beobachten, wie die „Formel“, der gerade ihre Aufmerksamkeit gilt, auf eine für sie typische Weise mit dem Kontext links und rechts der Formel zusammenhängt. Sie können beobachten, wie „umfassendere“ Muster entstehen und gezielt wahrnehmen, an welchen Stellen Variablen aufscheinen, die bei anderen Gelegenheiten mit anderen Inhalten auftreten werden. Sie erkennen, wie durch Inhalte Kohärenz geschaffen wird und wie diese Kohärenz mit der kohäsionsbildenden Form, mit den *Formeln* („formulaic language“) zusammenhängt.

## 7. Kostruktionen erkennen, erfahren, lernen

Es geht um eine propädeutische Hinführung der Lernenden zur Sprache, und zwar nicht nur um das Erkennen von formalen Mustern, sondern von Funktionstypen, die als Muster in Erscheinung treten. Die idiomatische Kompetenz des muttersprachlichen Sprechers besteht darin, keine Konstruktionen zu verwenden, die nicht auch von den anderen Teilnehmern der Sprachgemeinschaft verwendet werden. Aus der Gesamtmenge aller grammatikalisch möglichen werden nur die idiomatischen beachtet und gelernt. Das führt dazu, dass einige Konstruktionen mit weit größerer Wahrscheinlichkeit gebraucht werden und deshalb öfter in Alltagsgesprächen aufscheinen werden, als andere. Beim normalen Umgang mit den Mitgliedern der Sprachgemeinschaft, kommt der Lernende also oft mit den „usuellen", das heißt, den tatsächlich gebrauchten Konstruktionen in Kontakt und braucht sich nicht darum zu kümmern, eventuell mögliche Konkurrenzformen explizit auszuschließen[10].
Unsere Schulbücher sind aber voller sprachlicher Ausdrücke, die von der „Grammatik" generiert wurden, aber im Leben der Menschen keinen Ort haben. In der Schule haben sich Nischen der Nicht-Idiomatizität gebildet, von denen sich die vom Usus abweichenden Formen wie Viren ausbreiten, wenn man ihrer Verbreitung nicht entgegenwirkt. Südtirol, das mehrsprachige Land, spielt hier eine Vorreiterrolle, wo manche Probleme früher, konzentrierter und übersichtlicher in Erscheinung treten als anderswo. In Bayern und in Österreich finden wir den idiomatisch abweichenden Sprachgebrauch in Bereichen der Schriftlichkeit ebenso wie in Südtirol, aber als Einzelfälle, die nicht systematisch wahrgenommen werden. Bei uns erkennt der geschulte Beobachter ein und dasselbe Phänomen bei Sprechern des Deutschen als Erst- und als Zweitsprache dazu noch bei Italienisch als Zweitsprache. Dazu kommt das neue Phänomen der ganz und gar unidiomatischen neu-englischen Kommunikationssprache, die von Muttersprachlern des Englischen beinahe wie eine Fremdsprache gelernt werden muss und in der sich viele unserer globalisierten Mitbürger trotz der fehlenden *native-like selection* so wohl fühlen.

In dieser Situation einer zunehmenden sprachlichen Veränderung – ich würde von wachsender „Verunsicherung" sprechen, wenn man mir den erhobenen Zeigefinger durchgehen lässt – ist unser Projekt „Korpus Südtirol" entstanden, das einen Schwerpunkt in didaktischen Anwendungen hat. Die drei Schulsprachen Deutsch, Italienisch und Englisch sind gleichwertig präsent, sodass die ersten Erfahrungen mit der Ausrichtung auf Idiomatizität genau dort er-

10 Belica, in diesem Band, Steyer 2004.

lebt werden können, wo der einzelne Lernende seine ganz eigene persönliche Disponibilität mitbringt. Immer wieder haben wir erlebt, dass Studierende, denen die Korpusinstrumente an Hand deutscher Beispiele präsentiert wurden, spontan gefragt haben: „Gibt es das auch für unsere Muttersprache Italienisch?“
Der mehrsprachige Zugang ist ein entscheidendes Merkmal unseres Projekts, ein Fixpunkt, der bei jeder neuen Anwendungen aufs Neue seinen Wert erweist. Die Lernenden erfahren den Umgang mit Korpora sogleich als Anleitung zu einer „Haltung“ gegenüber den Sprachen, nicht als Instruktion, die auf das Erlernen bestimmter phraseologischer Ausdrücke abzielt.
Dass Korpus-Recherchen auch im Anfängerunterricht produktiv sein können, möchte ich an einem italienischen Beispiel zeigen. Eine deutsche Kollegin, die gerade begonnen hat, Italienisch zu lernen, erhält in der zweiten Stunde einen hochkarätigen Text aus der Kulturbeilage einer Tageszeitung, von dem sie auf den ersten Blick das Bild und die zwei Überschriften wahrnimmt:

Il dono della libertà
Il miracolo dell’identità

Der zweite Text mit seinen Internationalismen ist ohne Probleme zu verstehen, beim ersten denkt sie – eine typische *lectio difficilior* – an die Tücken des Begriffs Freiheit und schreibt dem unbekannten “dono” spontan negative Konnotationen zu.

Page 1 of 251 Go   Next | Last

dalla mia infermità , grazie Signore per il **dono** ***della*** tuo Spirito Santo nella mia vita
visita di funzionari cinesi in Sud Africa Il **dono** ***della*** sintesi chi sacrifica la libertà
più un' utopia ma noi cristiani abbiamo il **dono** ***della*** speranza partendo da ciò che già
retroguardia e l' amore divino ; portano con se il **dono** ***della*** guarigione conferitogli dalla loro
equiparare la grazia soprannaturale con i **doni** ***della*** natura , significa violentare il
tutti gli esseri : quando non possiedono il **dono** ***della*** trasformazione , col tempo sprofondano
allora non aveva mai pensato di desiderare il **dono** ***della*** trasformazione . Più di una scimmia
Gangi , rotta , è stata riappiccicata . **Dono** ***della*** fam. Vigo Pennisi di Gallodoro .
da poteri straordinari alcuni avevano il **dono** ***della*** guarigione , altri potevano accendere
Rossi " . Stato di conservazione : buono . **Dono** ***della*** famiglia Rossi di Acicatena . Inv
di conservazione . restaurato nel 1988 . **Dono** ***della*** Banca Popolare Santa Venera di Acireale
superficie del colore è alquanto inaridita . **Dono** ***della*** famiglia Romeo Marone . Inv . 37
x 46 . Stato di conservazione : buono . **Dono** ***della*** Società di S. Vincenzo dei Paoli
tentoni , nella loro ricerca chiedono a Dio il **dono** ***della*** pace e si impegnano affinché le loro
da Gerusalemme ( Lc 24 , 47 - 49 ) , nel **dono** ***della*** pace divina messianica e nell' investitura
universale . " Il santo non è qualcuno che ha il **dono** ***della*** profezia o quello di fare i miracoli
antropologia negativa per chi non disponga del **dono** ***della*** grazia può introdurre a un' interpretazione
forte richiamo , cioè : chi ha avuto questo **dono** ***della*** grazia e , possiamo dire , il dono
dono della grazia e , possiamo dire , il **dono** ***della*** fede , deve cercare – senza mai forzare
quella chiusura del sacrificio per mezzo del **dono** ***della*** passione e chiamano i credenti a

Bei der Suche mit der Teilformel „dono della“ stößt sie auf viel Unverständliches aber auch auf deutliche Indizien, die gegen ihre spontan erstellte Hypothese sprechen – eine Hypothese, über die sie sich nicht einmal Rechenschaft abgelegt hat. Ich gehe noch weiter. Die Lernende hat wahrscheinlich überhaupt nicht nach den Konnotationen von „dono“ gesucht, aber wegen ihrer fehlenden Sprachkenntnisse blieb alles Sprachliche ausgeblendet außer eben den stichwortartig eingeblendeten verständlichen Ausdrücken, die von ihr als „Werte“ wahrgenommen wurden:

Spirito Santo, natura, pace, grazia, fede, passione
famiglia Rossi, famiglia Romeo Marone
Banca Popolare, Società di S. Vincenzo dei Paoli

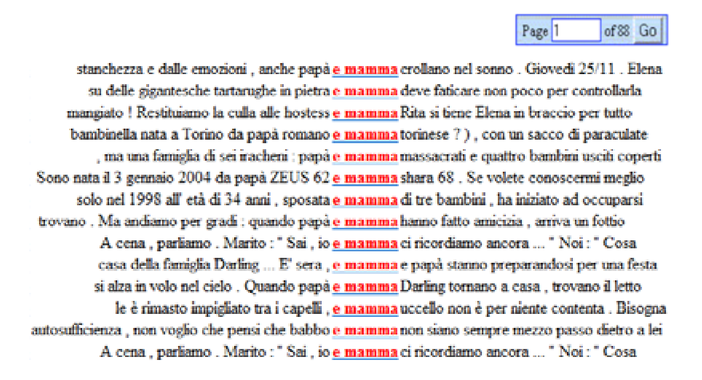
Sie hat in der KWIC-Liste etwas gefunden, ohne danach gesucht zu haben. Das ist das *Serendipity*-Prinzip der Korpussuche. Dass sie gerade Konnotationen wahrnimmt, ist kein Zufall, denn ihr spontanes Verständnis des Titels war ja stark von der Hypothese geprägt, „dono“ sei von negativen Konnotationen bestimmt. Sie hat also spontan aus der KWIC-Liste gezielt solche Belege kognitiv verarbeitet, die als Verifizierung/Falsifizierung ihrer Hypothese zu gebrauchen waren. Und kaum war sie zu einem Ergebnis gekommen, da war auch schon die hypothetische Zuordnung einer Bedeutung da. Sie identifizierte das italiensche „dono“ über das französische „don“ als „Geschenk“.
Dass sie überhaupt auf Ihre Französischkenntnisse zurückgegriffen hat und nicht, wie zu Beginn der Arbeit, ausschließlich auf das übersetzende Deuten ins Deutsche hinüber fixiert war, ist der Suchstrategie zu verdanken, die sie – über die Suche nach der Konnotation des ihr unbekannten Ausdrucks – von der Sprache weg in den Bereich der „Welt“, ihrer Sprachkenntnisse und ihres Weltwissens geführt hat.

Auch kleine, elementare Schritte können zu fruchtbaren Einsichten führen. Stellen wir uns einen Null-Anfänger vor, dem die Aufgabe zufällt, mit demselben Text (d.h. mit den Subtexten) den lexikalischen Bereich der Familie zu erarbeiten und als erstes die fehlenden Begriffe in der folgenden Bildunterschrift zu suchen.

**Amore di mamma.** Tabu è Ashima Ganguli, moglie di Ashoke e madre di Gogol. Nonostante la nostalgia dell’India decide di rimanere negli Usa per il bene dei figli.

Die Suche selbst soll ohne jede Instruktion von den Lernenden selbstständig definiert werden. Dazu dürfen allerdings aus methodologischen Gründen nur sprachliche Elemente verwendet werden, die zuvor bereits als „comprehended input“ erfasst worden sind.
Grundlage für die Arbeit ist die Annahme, dass im Artikel relevante Informationen zum Film gegeben werden, die das Verständnis der Bildunterschrift erleichtern (bzw. bestätigen) können.
Es werden die Konstruktionen *di mamma / di Ashoke / di Gogol* erkannt, wodurch die mit diesen Ausdrücken verknüpften Elemente *amore*, *moglie* und *madre* differenziert wahrgenommen werden können. Lässt man sich von den fremden Eigennamen nicht irritieren, dann werden Ausdrücke transparent, die sich zu einem Ganzen formen, wobei als isoliertes Element die Konjunktion „e“ erkannt wird, die im Kontext als Äquivalent zur deutschen Konjunktion „und“ erscheint.
Die Konjunktion ist damit als funktional eindeutig bestimmtes sprachliches Element verwendbar für die Definition der Recherche.
Gesucht werden soll nach dem männlichen Pendant zu „mamma“, zu „moglie“ und zu „madre“. Voraussetzung für die Suche ist die Annahme, dass „Vater und Mutter“ als feste Formeln belegt sind und das fehlende Element unmittelbar vor oder nach der Konjunktion „und“ aufscheinen wird. Ja, wir haben auch die Vorerwartung, dass die Formel hierarchisch gerichtet, zuerst den Mann anführt. Somit suchen wir nach:

e mamma


Page 1 of 88 Go

stanchezza e dalle emozioni , anche papà **e mamma** crollano nel sonno . Giovedì 25/11 . Elena
su delle gigantesche tartarughe in pietra **e mamma** deve faticare non poco per controllarla
mangiato ! Restituiamo la culla alle hostess **e mamma** Rita si tiene Elena in braccio per tutto
bambinella nata a Torino da papà romano **e mamma** torinese ? ) , con un sacco di paraculate
, ma una famiglia di sei iracheni : papà **e mamma** massacrati e quattro bambini usciti coperti
Sono nata il 3 gennaio 2004 da papà ZEUS 62 **e mamma** shara 68 . Se volete conoscermi meglio
solo nel 1998 all' età di 34 anni , sposata **e mamma** di tre bambini , ha iniziato ad occuparsi
trovano . Ma andiamo per gradi : quando papà **e mamma** hanno fatto amicizia , arriva un fottio
A cena , parliamo . Marito : " Sai , io **e mamma** ci ricordiamo ancora ... " Noi : " Cosa
casa della famiglia Darling ... E' sera , **e mamma** e papà stanno preparandosi per una festa
si alza in volo nel cielo . Quando papà **e mamma** Darling tornano a casa , trovano il letto
le è rimasto impigliato tra i capelli , **e mamma** uccello non è per niente contenta . Bisogna
autosufficienza , non voglio che pensi che babbo **e mamma** non siano sempre mezzo passo dietro a lei
A cena , parliamo . Marito : " Sai , io **e mamma** ci ricordiamo ancora ... " Noi : " Cosa

und finden sogleich einen Beleg mit „papà“ an der vermuteten Stelle, dazu aber eine Fülle von weiteren sprachlichen Informationen über Wörter im Kon-

text, die Pluralformen sowie Details der Morphologie, die alle in klar definierten Konstruktionen funktional in Erscheinung treten und mit den dazu gelieferten Parallelstellen verifiziert, zugleich wiederholt werden können.
Da „papà" in Zeile 10 auch rechts von „mamma" überliefert ist, liegt es nahe, auch diese Variante sogleich zu überprüfen. Man erhält dabei einen Einblick in die Kraft der Konventionalität, denn die Formel „mamma e" ist nicht nur um den Faktor fünf häufiger dokumentiert, sondern, zumindest im Singular, äußerst homogen als Formel „mamma e papà" überliefert.

Page 1 of 443 Go

| | | |
|---:|:---:|:---|
| . E sento il profumo di tutti i ragù che | **mamme e** | nonne e zie hanno cucinato , per condire |
| bimbi possono crescere con i Nonni , con | **Mamma e** | Papà , con la nuova moglie di Papà , |
| scelta dell' associazione Tornati in Italia | **Mamma e** | Papà , telefonano a questa associazione |
| arriva il giorno di partire per Albenga , | **Mamma e** | Papà non stanno più nella pelle perché |
| Agosto 1997 , alla nascita pesava 2,7 Kg. | **Mamma e** | Papà decisero che quello era " il loro |
| foto arriva il 10 giugno 1999 e finalmente | **Mamma e** | Papà conoscono il loro bimbo ... che |
| preparare il passaporto . Dopo 5 mesi più o meno | **Mamma e** | Papà sarebbero dovuti andarlo a prendere |
| prendere . A luglio Kristian telefona a casa di | **Mamma e** | Papà che subito si allarma : forse qualcosa |
| " piccoli cucciolotti " . Nel frattempo | **Mamma e** | Papà decidono quale sarà la loro cameretta |
| molto più felici . Entreranno nel lettone di | **Mamma e** | Papà ? Andranno a svegliarli con un bacio |
| possibile . Manca poco cucciolotti , dopo | **Mamma e** | Papà non vi lasceranno più !! A metà |
| torna con il foglio firmato e timbrato ! | **Mamma e** | Papà non credono ai loro occhi ! Mamma |
| arrivati ! Purtroppo così non è stato . | **Mamma e** | Papà sono potuti partire solo il 14 maggio |
| rifare i passaporti ! ... sarà poi vero ? ) . | **Mamma e** | Papà non ce la fanno più ad aspettare |
| sull' aereo ! Alla fine l' aereo decolla . | **Mamma e** | Papà non stanno nella pelle : domani |
| cerca di spiegargli che quei signori sono | **Mamma e** | Papà : sono venuti a prendere lui e la |
| dalla valigia ( posizionata sul sedile tra | **Mamma e** | Papà ) tutto quello che può arraffare |
| tempo si rendono conto che " quei due " | **Mamma e** | Papà solo lì solo per loro , per aiutarli |
| forse il più difficile da gestire : ... ma | **Mamma e** | Papà si devono abituare . Ora sono genitori |
| un bacio a questi meravigliosi bambini da | **Mamma e** | Papà : grazie di esistere !!! ... e un |

Wer an diesem Punkt die Varianten im Plural erkunden möchte, muss nur die Pluralform, die in der ersten Zeile belegt ist, „mamme e" in das Phrase-Fenster eintippen und schon erscheinen die „nonne e zie", die „bambini" aber auch die „papà" im Plural.
Und weiter mit

e moglie
e madre

und den Umkehrungen. Schon die kursorische Analyse der ersten Seite der jeweiligen Suchergebnisse ergibt eine Fülle von brauchbaren Kontexten, Formeln, Konstruktionen, Gruppen und den jeweiligen morphologischen Merkmalen, die wie auf einem Silbertablett präsentiert erscheinen. Dass die

relevanten Beispiel aus einer Fülle von unverständlichen oder weniger interessanten aktiv ausgewählt werden müssen, verstärkt den Effekt der gerichteten Aufmerksamkeit. So werden nicht nur lexikalische Einheiten gelernt, dazu eine erste syntaktische Struktur von hohem Gebrauchswert, sondern es werden zugleich Erwartungen aufgebaut, größere Zusammenhänge kommen ins Blickfeld und idiomatische Eigenheiten der Zusammenstellungen von Elementen werden von Anfang an mitgelernt.

## 8. Die Freude am Lernen und am Entdecken

Aus diesen Erfahrungen kann zwar keine stringente Folgerung abgeleitet werden, aber immerhin eine neue Frage. Könnte es der Fall sein, dass erfolgreiche Sprachlernprozesse – auch – damit zu tun haben, dass es den erwachsenen Lernenden gelingt, als Grundlage für den Gebrauch der neuen Sprache feste Fügungen ins Gedächtnis aufzunehmen und nicht etwa die gleiche Anzahl isolierter Wörter, die erst später (nach dem *Just-In-Case*-Prinzip) unter Anwendung der gleichzeitig gelernten Grammatikregeln „kreativ" zu gebrauchsfertigen Einheiten zusammengeführt werden?
Was, wenn Lernende hingegen in den ersten zwei Wochen Unterricht einige hundert dieser „Konstruktionen" lernten? Jede „formelhafte" Mehr-Wort-Verbindung wird im Gedächtnis gespeichert als notwendiger Ausdruck einer ganz speziellen Situation, deren Einbettung in das enzyklopädische Wissen des Lesers durch Eigennamen, historische Angaben und dergleichen mehr gewährleistet ist. Jede Formel wird als sprachlicher Ausdruck, also als Form, wahrgenommen und kann zugleich holistisch einem Moment des täglichen Lebens, einer *tranche de vie* zugeordnet werden.
Diese Beschreibung der „funktionierenden" Sprache führt zu einem der noch ungelösten Probleme des Wortschatzerwerbs beim Erlernen einer Fremdsprache. Wie kommt der analytische Anteil des Lernens ins Spiel, wenn der holistische stark bevorzugt und sogar antizipiert werden soll? Die Übungsanordnung mit Korpusinstrumenten, die bei holistisch verarbeiteten Konstruktionen ansetzt, ist der Versuch, auf diese Frage eine Antwort zu geben. Bei der umgekehrten Lernstrategie, bei der Strategie nämlich, mit dem analytischen Lernen zu beginnen und auf diese Weise eine Kompetenz anzustreben, die Schritt für Schritt erweitert wird, erscheint der idiomatisch adäquate Sprachgebrauch überhaupt nicht am Horizont des Möglichen. Ein Meer von „Grammatikfehlern" trennt die Lernenden von diesem Ziel, das – ebenso wie die prosodische, die stilistische, die soziolinguistische und die textlinguistische Angemessenheit – aus den Curricula ausgeklammert bleibt.

Wenn hingegen die idiomatische Sprache an den Anfang des Lernens gestellt wird, dann kann die Frage nach der Verbindung von holistischem und analytischem Vorgehen beim Lernen nicht ungestellt bleiben und verlangt nach einer Antwort. Allein dieser Zwang stellt schon einen Gewinn dar, den wir nicht ungenutzt lassen sollten.

Der von verschiedenen Autoren vorgeschlagene Weg, die holistisch gelernten Formeln in einem Folgeschritt gleichsam aufzuknacken und dem analytischen Lernen als Lernstoff weiterzureichen, wird von Alison Wray entschieden abgelehnt. Holistische Formen der Sprache haben es an sich, dass sie beim „normalen" Gebrauch durch kompetente Sprecher eben gerade nicht analytisch behandelt werden. Die Idiomatizität stellt gleichsam eine Schutzhülle gegen das Analysieren der Einzelteile dar. Beim Lernen in der Schule sollte man nicht gegen dieses funktionale Prinzip verstoßen, so Wray, obwohl es nahe liegt, einen Kompromiss zu suchen und wie einer der von Wray zitierten Autoren „to accept the contradiction between his goal, which is to bring learners to a mastery of native-like expression, and his means of achieving this, which entails non-native-like processing behaviour and a tendency to overgeneralize". (Wray 2000: 484) Aber damit ist noch nicht das letzte Wort gesprochen. Im Gegensatz zu Übungsformen, bei denen idiomatische Formeln isoliert in den Unterricht eingebracht werden, erscheinen bei der selbst-verantworteten Korpus-Suche die Formeln in einer Vielzahl von Belegen. Schon die Suche nach verwertbaren, prototypisch erfahrbaren Belegen führt zu einem Priming-Effekt in Hinblick auf den gleich bleibenden Bestandteil der Belege, den Suchbegriff, der durch Fettdruck hervorgehoben ist und sich daher leicht einprägt. Der Suchbegriff wird im Lauf der Arbeit zwischen fünf und zehn Mal bewusst wahrgenommen und hat somit eine hervorragende Chance, schon im Lauf dieses ersten Arbeitsschrittes ins Langzeitgedächtnis übernommen zu werden.

Eingespeichert wird dabei die vollständige Konstruktion in ihrer Synthese aus Form und Funktion. Dazu kommen Erinnerungen an die thematische Umgebung, an Namen, Orte, Institutionen, Internationalismen, sowie die Funktion der Formel im jeweiligen Kontext. Die „Bedeutung" als gesonderter Bestandteil spielt dabei noch keine Rolle. Sie kommt erst in einem zweiten Schritt, nach der Wahl des „Prototypen", der zum Feinverständnis des Ausdrucks führt, zum bereits gespeicherten Ausdruck hinzu.

Während der mit der Suche verbundenen unterschiedlichen Arbeitsschritte muss sich der Benutzer der Query-Software darauf besinnen, dass er die Suche nicht wegen der Suche selbst durchgeführt hat, sondern in der Absicht, mit Hilfe der Suchergebnisse einen Schritt weiter auf dem Weg des Erlernens

der Sprache zu tun. Er muss dafür den Suchausdruck fokussieren und gezielt wahrnehmen, sollte ihn vielleicht leise mitsprechen, um die phonetische Grundlage für die Verarbeitung im Arbeitsgedächtnis zu schaffen. Er soll den Mehr-Wort-Ausdruck in seiner Gesamtheit wahrnehmen und sich auf weitere Suchen vorbereiten, die der Stärkung der Sprachaufmerksamkeit dienen.
Durch den oftmaligen Gebrauch bilden sich Erwartungen und Suchstrategien heraus, die ich jetzt als mögliches Lernziel beschreiben möchte. Worin besteht die Kompetenz, die bei der Durchsuchung einer umfassenden Sammlung von authentischen Texten nach einzelnen Mustern so gute Dienste leistet? Anders gefragt. Wonach sucht der erfahrene Nutzer? Das heißt, ein Nutzer, der diese Suchen schon sehr oft gemacht hat, tausende von Malen?
Ich verwende die hier beschriebenen Instrumente und die hier vorgestellte Methode beinahe täglich in den Sprachen Deutsch, Italienisch und Englisch. In der Analyse des Selbstversuches erkenne ich, dass ich spontan nach Indizien für den Gebrauchswert der Redewendung suche, deren genaue Bedeutung ich mit Hilfe dieses Instruments erkunden möchte. Ich suche also nicht nach der „Bedeutung“, wie sie sich etwa als „Übersetzung“ des unbekannten Ausdrucks zeigen könnte. Tausendfaches Suchen hat mich geübt gemacht im raschen Lesen, im Überfliegen der Kurztexte, im Verzicht auf eine rasche Antwort auf Grund der ersten Zeilen zu Gunsten eines Bildes, das sich nach dem Vergleich mit den nachfolgenden Kurztexten ergeben könnte. Das „Aufschieben des Verstehens“, zu dem nach Nietzsche eine größere „Moralität“ vonnöten ist, als beim raschen Reagieren auf einen sprachlichen Reiz. Warum habe ich nicht ein Wörterbuch gewählt? Vielleicht, weil ich gerade diesen Prozess des Vergleichens, des Erprobens von Hypothesen suche und wünsche. Oder, weil ich mit dieser Art von Suche gute Erfahrungen gemacht habe. Oder einfach, weil ich mehr Informationen brauche als die, die ich von einem Wörterbuch erwarte, nämlich Informationen über Kontexte, Textsorten, in denen die Redewendung überliefert ist, Konnotationen, die bei ihrem Gebrauch mitschwingen und die ihren Gebrauch einschränken.
Dazu kommt ein Lustgefühl, das im 19. Jahrhundert als „Flanieren“ beschrieben wurde, und das im Umfeld der Korpus-Recherchen *Serendipity*-Erfahrung genannt wird: Das glückhafte Finden von etwas, wonach man eigentlich gar nicht gesucht hat.

## 9. Auf den Spuren der Interjektion *In Gottes Namen*

Auf allen Stufen der Lerner- und Interimskompetenz können Korpusrecherchen einen Beitrag zum Aufbau und Wachsen der Sprachkompetenz leisten.

Die höchste Stufe im Rahmen des Themas Sprache und Schule, stellt der Muttersprachunterricht am Gymnasium dar. Hier gibt es junge Menschen, die für metasprachliche Aspekte der eigenen Kompetenz und für formale wie funktionale Einsichten in das Funktionieren der eigenen Sprache empfänglich sind. Die Zeit haben, sprachliche Feinheiten an literarischen Texten zu überprüfen und zu erfahren. Diese Chance muss im Unterricht genutzt werden.
Die Stärkung der Muttersprachkompetenz in der Südtiroler Schule hat zwei ganz konkrete Aufgaben. Es gibt neuralgische Punkte im Kontakt Dialekt-Hochsprache und kritische Momente der Sprachkompetenz im Bereich der Bildungssprache, wo sich Schwachstellen oft bei phraseologischen Ausdrücken zeigen. Solche Schwachstellen sollen nicht überbetont und explizit thematisiert werden, aber sie sollen auch nicht verdrängt und vergessen werden. Ich möchte hier eine Korpus-Analyse aufgreifen, die seit Jahren zum Standardrepertoire meines Unterrichts in mehrsprachigen Klassen gehört. Es ist dies ein sehr aufschlussreiches Beispiel für die enge Verflechtung der Leseakte mit vorgefassten Frames und den Ergebnissen der Korpussuche mit interpretationsbedürftigen Ergebnissen.
Die notwendige Interpretation mit den Gefahren des Überinterpretierens, des falschen Interpretierens und des Missverstehens kann auf Grund des Verfahrens, das auf einer Vielzahl von Daten basiert, sehr effizient in ein didaktisches Verfahren pluraler Leseakte überführt werden. Die Einsichten über Defizite und Schwachstellen der Sprachkompetenz der Schülerinnen können unausgesprochen in diese Arbeit einfließen und auf Lernstrategien abzielen, die genau diese Defizite beheben helfen ohne sie plakativ hervorzuheben.

Ausgangspunkt ist ein Text von Hans Magnus Enzensberger. In Enzensbergers *Der Untergang der Titanic*, einem Gedichtzyklus aus 33 Gesängen, finden sich 16, zum Teil stark fachsprachlich geprägte, Einschübe oder Exkurse, wie sie in Enzensbergers Lyrik so typisch sind. Der zweite dieser Einschübe trägt den Titel *Verlustanzeige* (Enzensberger 1978: 65) und setzt stereotype sprachliche Elemente in Verbindung zu Interjektionen.

**Verlustanzeige**
Die Haare verlieren, die Nerven,
versteht ihr, die kostbare Zeit,
auf verlorenem Posten an Höhe
verlieren, an Glanz, ich bedaure,
macht nichts, nach Punkten,
unterbrecht mich nicht, Blut

verlieren, Vater und Mutter,
das in Heidelberg verlorene Herz,
ohne mit der Wimper zu zucken,
noch einmal verlieren, den Reiz
der Neuheit, Schwamm drüber,
die bürgerlichen Ehrenrechte, aha,
den Kopf, in Gottesnamen, den Kopf,
wenn es unbedingt sein muß,
das verlorene Paradies, meinetwegen,
den Arbeitsplatz, den Verlorenen Sohn,
das Gesicht, auch das noch,
einen Backenzahn, zwei Weltkriege,
drei Kilo Übergewicht verlieren,
verlieren, immer nur verlieren, auch
die längst verlorenen Illusionen,
na wenn schon, kein Wort
über die verlorene Liebesmüh,
aber woher denn, das Augenlicht
aus den Augen, die Unschuld
verlieren, schade, den Hausschlüssel,
schade, sich, gedankenverloren,
in der Menge verlieren,
unterbrecht mich nicht,
den Verstand, den letzten Heller,
sei's drum, gleich bin ich fertig,
die Fassung, Hopfen und Malz,
alles auf einmal verlieren,
wehe, sogar den Faden,
den Führerschein, und die Lust.

Auf den ersten Blick ist die Struktur des Textes einsichtig. Stereotype Formeln und feste Syntagmen mit dem Verbum „verlieren" werden zusammengestellt und durch sogenannte „sekundäre" Interjektionen oder Diskurs-Marker in der jeweiligen Sprechsituation verankert. Neu ist gerade die Kombination von idiomatischen Formeln und Interjektionen. Enzensbergers Text zwingt den Leser gleichsam, jede Interjektion in ihrer virtuellen Situation zu verstehen und daher eine passende Situation zur jeweiligen Aussage hinzuzudenken. Dadurch werden auch die Phraseologismen in ihrem situativen Gebrauchswert erfahrbar: „die Nerven verlieren", „kostbare Zeit verlieren". Ein kleines

„Frame-fülle-dich-Spiel“ kann beginnen, bei dem die Schülerinnen und Schüler Situationen suchen und (er)finden sollen, bei denen die einzelnen Frames in persönlich erfahrenen Aktualisierungen sprachlich realisiert werden.
Interjektionen entfalten ihre Funktion im Raum des Sprechaktes auf sehr unterschiedliche Weise. Sie verbinden kognitive und affektive Aspekte und können, in verschieden starker Gewichtung alle drei Funktionsklassen des Bühlerschen Modells betreffen, also sprecher-, adressatenbezogen oder stärker auf die Sachverhalte bezogen sein. Diese Variabilität kann beim Aktualisieren der Ausdrücke aus Enzensbergers Text nuanciert erfahren werden.
Die jeweils konkreten Sprechsituationen, die von Enzensbergers Textabschnitten verlangt werden, sind aber nicht vorgegeben, wie auch die kontextuelle Bedeutung der einzelnen Interjektionen des Textes einen weiten Spielraum an Sprachgebrauch abdecken kann. (Reisigl, 1999)
Zu diesen Erfahrungen mit dem Text kommt sodann die Feinanalyse der zentralen Stelle, die beim ersten Lesen wahrscheinlich gar nicht auffällt. Wie genau Enzensberger die Abfolge von Verlustanzeige und Einspruch der Stimme der Interjektionen komponiert, sieht man an der Abfolge der Ausdrücke „den Reiz der Neuheit verlieren – Schwamm drüber“ und „das verlorene Paradies – meinetwegen“; die Stimme bleibt also keineswegs unbeteiligt als Ausdruck stereotyper Kommentare. Das provozierende „meinetwegen“ als Antwort auf den Verlust des Paradieses dient als Signal für die persönliche Einstellung zu den Situationen, die der Text mit seinen rudimentären Sprachelementen evoziert. Dieses deutliche Signal steht nach drei Interjektionen, deren Funktionsbestimmung ebenfalls dem Leser überantwortet ist – nach einem enigmatischen „aha“, einem unvertrauten „in Gottesnamen“ und dem keineswegs harmlosen „wenn es unbedingt sein muss“:

die bürgerlichen Ehrenrechte, aha,
den Kopf, in Gottesnamen, den Kopf,
wenn es unbedingt sein muß,

Werfen wir zuerst einen Blick auf die Interjektion *in Gottesnamen*, die wohl nicht alle deutschen Sprecher im aktiven Wortschatz haben. Die Frage des Gebrauchs stellt sich also in der Form: in welchen Kontexten Sprecher, die *in Gottesnamen* überhaupt als Interjektion verwenden, den Ausdruck tatsächlich verwenden.
Es gibt mehrere, nebeneinander bestehende Gebrauchsformen von *in Gottes Namen*, die sich auch durch unterschiedliche prosodische Realisierungen des Syntagmas deutlich voneinander abgrenzen lassen. Eigentlich handelt es sich

um eine Art Homonymie von zwei klar getrennten Formen, von denen nur eine als sekundäre Interjektion gelten kann. Die Form mit voller Bewahrung der Wortsemantik von „Gott" und „Namen" erfüllt ja gerade nicht die Bedingung für die Zugehörigkeit zur Klasse der sekundären Interjektionen, die dadurch definiert ist, dass die Wortgruppe auf dem Weg zur Interjektion ihren ursprünglichen wörtlichen Sinn eingebüßt hat. Also sind die folgenden Beispiele aus dem Mannheimer Corpus nicht als Interjektionen zu werten:

Mit den Worten „Wir wollen **in Gottes Namen** gießen" läßt er den Guß beginnen.
Vorbild für den Funk **in Gottes Namen** ist Radio Maria Italien. Vor sieben Jahren gegründet, ist der Sender nach RAI 1 der zweitstärkste im Land.

Bei dieser Verwendung wird das Syntagma mit steigendem Melodieverlauf gesprochen und unterscheidet sich daher deutlich von den Fällen der Interjektion, die nicht mit dem Hauptton versehen ist und die durch eine fallende Intonationskontur und Deakzentuierung von „Namen" gekennzeichnet ist.
Vielen, aber keineswegs allen Sprechern des Deutschen sind diese beiden Formen vertraut. Werden Muttersprachler des Deutschen aufgefordert, im Kontext einer didaktischen Demonstration Beispiele für diese Interjektion anzugeben, dann geben sie in großer Mehrzahl zuerst ein Beispiel für die wörtliche Verwendung des Syntagmas. Das ist ein schönes Beispiel dafür, wie der Beobachter den Gegenstand seiner Beobachtung durch die Fragestellung beeinflusst und bereits vorentscheidet, denn die mündlich vorgebrachte Aufforderung, ein Beispiel anzugeben, versieht das Syntagma *in Gottes Namen* mit starker Betonung und steigender Intonationskontur, wie sie bei Zitaten üblich ist, und gibt dem Hörer das akustische Signal, nach einem Syntagma mit steigendem Ton zu suchen. Die Antwort wird vom Priming der Frage bestimmt, die bereits ein bestimmtes prosodisches Muster vorgibt.
In vielen Fällen, die mir untergekommen sind, stutzt die Informantin aber gleich nach ihrem Beispielsatz und sucht nach einem weiteren Beispiel, das ihrer Intuition über den Gebrauch der Interjektion „in Gottes Namen" eher entspricht als der erste Ausdruck, den sie gerade als Beispiel angegeben hat. Die zweite Aktualisierung gibt dann ein Beispiel für die Interjektion mit der ihr eigenen Semantik und der für sie typischen prosodischen Realisierung.
Die ersten authentischen Beispiele, die ich einer Kollegin abverlangt habe, zeigen die zwei unterschiedlichen prosodischen Realisierungen des Syntagmas, beim ersten Beispiel mit starker Betonung, beim zweiten Beispiel hingegen erhält das Verb den Haupton und die Interjektion ist deakzentuiert.

Lass mich in Gottes Namen in Frieden!
Dann geh' halt in Gottesnamen.

Die Kollegin hat die zwei Beispiele – auf meine Bitte hin – aus einer für sie erfahrbaren vorgestellten Situation heraus laut gesprochen, und nach einigem Zögern waren die zwei Intonationsmuster deutlich differenziert.
Auch der Muttersprachler hat also Anlass, sich seiner Sprachkompetenz zu vergewissern, und die experimentellen Bedingungen stellen bei dieser Kontrolle schon eine erste, nicht zu übersehende Hürde dar.
Bei dieser gesteuerten und kontrollierten Reaktion auf eine Aufgabe kann man gleichsam am eigenen Leib die Erfahrung sprachlicher Kommunikation in Form von Prototypen erleben. Wir haben keine genaue Vorstellung von der Bedeutung des Ausdrucks *in Gottes Namen*, aber wir haben eine Vorstellung von Situationen, in denen der Ausdruck verwendet werden kann, wenn ihn jemand verwendet. Und wir erkennen auch den semantisch vollen Ausdruck als solchen. Die Vorstellung von erinnerten oder simulierten Situationen ist individuellen Schwankungen unterworfen, die bei diesem Beispiel einen großen Spielraum abdecken. Aber Sprecher, denen beide Formen vertraut sind, haben das Gefühl, dass *in Gottesnamen* in den ihnen gerade präsentierten Beispielen richtig verwendet wird.
Die beim normalen Sprachgebrauch unzugängliche Erinnerung an die Interjektion wird durch den literarischen Text aus dem Gedächtnis hervorgerufen und wieder in den aktiven Wortschatz aufgenommen. Dieser Prozess kann durch die gezielte Suche nach Beispielen gefördert werden.
So viel zu den Studierenden, denen die Interjektion bekannt ist. Es gibt Dialektsprecher, denen die prosodische Gestaltung mit dem Akzent auf NAmen so unvertraut ist, dass es ihnen schwer fällt, diese lautliche Form auch nur nachzusprechen. Spontan erhält immer GOTtes den Hauptton des Ausdrucks. Andere Studierende wiederum erlangen erst nach mehreren Hinweisen Zugang zur Erinnerung an die Form mit Betonung auf NAmen. Und es sind immer wieder Hinweise auf konkrete Sprechsituationen, der konkrete Hinweis auf die Großmutter, den Großvater oder andere ältere Mitglieder der Familie, die zum Aha-Erlebnis führen. Ja, sie erinnern sich an diese Variante und sind sogleich imstande, die Funktion des Ausdrucks zu rekonstruieren: „Wenn jemand etwas ungern tut oder widerwillig ein Zugeständnis macht...".
Die Erfahrungen mit spontan geforderten Beispielsätzen zu *in Gottes Namen* zeigen, wie schwierig es ist, den eigenen Sprachgebrauch in einer kontextfreien Situation zu rekonstruieren. Man erkennt rasch, dass es nicht ganz einfach ist, authentische Beispiele zu finden und sie situationsadäquat laut einer Gruppe

von Zuhörern vorzusprechen. Diese Erfahrungen legen es nahe, den Lehrpersonen Anregungen für die Stundenvorbereitung zu geben. Korpusrecherchen bieten sich als gezielte Materialsammlung dafür an. (vgl. McCarthy, 2004)
Was geschieht nun beim Lesen von Enzensbergers Text? Die Leser „müssen" die Interjektion verstehen, sie sind daher „gezwungen", eine Situation zu konstruieren, in der die Verwendung der Interjektion sinnvoll erscheint. Was für eine Situation ist das, und welcher Prototyp von *in Gottes Namen* wird dabei aktiviert? Diese Entscheidung steht und fällt bereits mit dem Verständnis des idiomatischen Ausdrucks „den Kopf verlieren", der den engeren Kontext der Stelle darstellt. Das Verständnis dieses zweideutigen Ausdrucks wird aber ebenfalls vom Kontext gesteuert, so dass wir hier einen Paradefall für das von Nietzsche geforderte „sorgfältige, rücksichtsvolle Lesen" haben. Ein Lesen mit „offen gelassenen Türen", Türen, Zugängen zum Verstehen, die nicht auf einen Schlag zugeworfen werden, sobald wir eine erste Hypothese des Verstehens gebildet haben (vgl. Drumbl, 2003b, 2007b). Die gegenseitige Bedingung der beiden Ausdrücke verlangt nach einem Ausgleich der jeweiligen *constraints* gegenüber dem Kontext. Die Bedeutungen müssen für beide Fälle und in beide Richtungen hin passen. Die Bedeutung der Zeile wird gleichsam unter dem Druck der *constraints* ausgehandelt.

den Kopf verlieren (unüberlegtes Handeln)
den Kopf verlieren (getötet werden)
in Gottes Namen (den Namen Gottes anrufend)
in Gottes Namen (Interjektion: unter Druck gegebenes Einverständnis)

Dieser Prozess des Erprobens von kompatiblen Bedeutungen, der hier schematisch vorgeführt wird, stellt den Kern der Text- und Diskurskompetenz in der Muttersprache dar.
Lernen können wir diese Kompetenz als Sprecher der Zweitsprache – auch – durch Beobachtung von gelungenen Fällen ihrer Realisierung in den jeweils spezifischen Kontexten. Für jeden denkbaren Sprachgebrauch gibt es Präzedenzfälle, an denen man die gerade anstehende Aufgabe messen kann.
Das „Glücken" der eigenen Sprechakte ist eingebettet in Sprechakte anderer Teilnehmer der Sprachgemeinschaft, geglückte, wie misslungene Sprechakte, von denen wir – in Ausnahmemomenten gesteigerter Aufmerksamkeit – einige wenige wahrnehmen. Die Begegnungen mit literarischen Texten sind solche Ausnahmefälle.
Enzensberger hat den Text geradezu auf die Interjektion „in Gottesnamen" hin komponiert. Die polemisch aktuelle Anklage, einem alten Ausdruck an-

vertraut, der nur mehr wenigen Mitgliedern der Sprachgemeinschaft im aktiven Wortschatz geläufig ist.
Enzensberger ist die Interjektion seit seinen Studienjahren vertraut und es scheint, als habe er ein halbes Leben lang darauf gewartet, ihr den gebührenden Platz in einem eigenen Werk geben zu können. Ambivalent als Anrufung höchster moralischer Instanzen in Momenten sprachlichen Handelns, das auf schändlichste Art und Weise den moralischen Grundlagen der eigenen Kultur widerspricht. Diese Qualität der Interjektion im Gedicht wird schlagartig klar, wenn man im Mannheimer Korpus auf folgenden Beleg aus dem Zürcher Tagesanzeiger stößt:

Deren heimliche Moral: Männer Gottes soll man nicht hinrichten. Die andern darf man, wenn es denn **in Gottes Namen** sein muss.

Das ist genau die Bedeutung der Interjektion in Enzensbergers Text. Enzensberger hat dem idiomatischen Ausdruck ein kleines Denkmal gesetzt, hat den Ausdruck auf ein Podest erhoben, wo er deutlich sichtbar, vom Kontext feinstimmig abgesichert, sein in der Alltagssprache tief verwurzeltes Potential zur Entfaltung bringt. Wer diesen Ausdruck in seiner eigenen Sprache kennt, weil er in den Situationen, in denen er funktional auftreten kann, tatsächlich verwendet wird, ist durchaus imstande, genau diese Funktionalität auch im literarischen Kontext wahrzunehmen. Den Leser trifft dann die volle Macht der unvorhergesehenen Verbindung der Interjektion, bei der die Anrufung Gottes mitwirkt, mit der Zustimmung zum Töten eines Menschen.
Konflikte spielen im Alltagsgebrauch der Interjektion eine große Rolle. Dazu einige Belege aus aktuellen und oft politisch brisanten Kontexten:

Gadamer: „Geht in irgendein anderes Studienfach oder, **in Gottes Namen**, werdet ein gebildeter Taxichauffeur!“
Es ist doch ein Schaden für unsere Demokratie, wenn ein Teil der Wohnbevölkerung abseits steht. Wenn man dieses Defizit nur mit einem Sonderangebot beheben kann, warum soll dann **in Gottes Namen** die doppelte Staatsbürgerschaft nicht ein solches Sonderangebot sein?
„Man muß doch mal vergleichen, was die Buche den Menschen bringt und was der geplante Neubau bringt“. Für Marianne Götze ein klarer Fall: „Fällt **in Gottes Namen** den Baum!“. „Ich bin dafür, daß die Buche stehen bleibt“, sagte dagegen Günter Ehrenfried klipp und klar.
Nun könnte man sagen, die CDU soll sich nicht so anstellen mit ihren Prinzipien und aus Opel in Gottes **Namen** dann eben eine Art zweiten Volkswagen machen. Das wirkliche Problem ist aber, dass nach wie vor niemand weiß, ob und wie Opel überhaupt zu helfen ist.
online/2009/14/opel-merkel_dwdsddc.xml

Der Kontext wird transparent im Moment, in dem die vertraute Redewendung als vertraut erkannt wird und zu vertrauten Kontexten in Beziehung gesetzt wird, die vom literarischen Kontext auch weit entfernt sein können.
In solchen Momenten kann ein Lernprozess stattfinden, der eine Wendung auf Dauer ins Gedächtnis einprägt. So etwas bewirkt die Stelle im Gedicht von Enzensberger. Er selbst hatte die Interjektion an einer literarisch auf höchstem Niveau markierten Stelle gefunden, und dazu in einem höchst unerwarteten, wenn nicht sogar unpassenden Kontext: „Nun gehe in Gottes Namen, du wunderlich liebe verfluchte Seele, fort, fort!" Eine Stelle bei Brentano, die Enzensberger seinerzeit in seine Brentano-Anthologie aufgenommen hat, und die ihm wohl schon als Student, während der Arbeit an seiner Doktorarbeit zur Poetik Brentanos aufgefallen war.
Es ist die Abschiedsszene zwischen dem Erzähler und der jungen Perdida:

„Du verwirrtes satanisches Geschöpf! Nun packe dich!" – „Sieh, der Morgenstern, gleich wie ein Held, will uns den Tag verkünden, leb wohl! Leb wohl! Was wirst du nun machen, wenn du die kleine schöne liebe verlorene nichtsnutzige Perdida nicht mehr hast?" – „Noch eins, Mädchen! Wer ist dann schuld an deiner verdammten Seele?" – „Ei, das Katholischwerden der verzwickten Dichterin, die sich hier für eine Gräfin ausgiebt, sie lügt und ziert und schraubt, da muß man es ja lernen, und doch hat sie mich weggejagt, weil ich keinen Sinn für höhere Kunst habe, aber sie war eigentlich eifersüchtig auf mich – ho, ho!" – „Nun gehe in Gottes Namen, du wunderlich liebe verfluchte Seele, fort, fort!" – und somit lief die Eidechse mit dem Hündchen und ihrem Raub singend ab: „Ich bin ja so geschwinde als ein Mädchen kann sein!" – Ich aber blieb zurück, und als ich das kleine Kreuz, das sie mir zurückgelassen hatte, ergriff, durchdrang mich eine namenlose Rührung, ich war fest entschlossen, Rom auf einige Zeit zu verlassen, wenige Kleider packte ich zusammen, steckte meinem Hausherrn einen Brief an die Türe, und begab mich zu meinem Freunde Bonascopa. (S. 247)

Spielerisch leicht findet die Interjektion bei Brentano ihren Ort. Aber ebenso natürlich und unaufdringlich erscheint sie an vielen Orten in literarischen Texten: von den Wahlverwandtschaften, „Nimm Ottilien, laß mir den Hauptmann, und in Gottes Namen sei der Versuch gemacht!" bis zum kaum vernehmbaren „in Gottes Namen", mit dem im „Rosenkavalier" die Marschallin Octavian ins Liebesglück mit Sophie entlässt.

* * *

Die Korpusrecherchen zu „in Gottes Namen“ haben paradigmatischen Charakter. Der Ausgangspunkt der Suche nach Belegen war ein prägnantes literarisches Beispiel, das die Interjektion „fragwürdig“ werden ließ. Die Belege aus der ersten Suche (im Mannheimer Korpus) brachten einige klar erkenntliche Parallelfälle zum Gebrauch bei Enzensberger, von Goethe bis zu Gadamer, mit dem aufschlussreichen Beleg, bei dem vom Erschießen von Gefangenen die Rede ist. Dazu kamen die Belege, die auf den ersten Blick als inkompatibel zur Bedeutung der Interjektion zu erkennen waren, mit der starken Betonung auf GOTtes und der nicht-phraseologischen Struktur der Mehr-Wort-Gruppe. Jemand beginnt etwas und ruft den Namen Gottes an.
Die Erkenntnis, dass in der Fülle der Belege zwei deutlich voneinander getrennte Muster aufscheinen, leitet den Blick des Suchenden sodann, die Belege nach den Kategorien der Zugehörigkeit zu diesen zwei Mustern zu ordnen. Da das Interesse der Interjektion gilt, wird die Gruppe der wörtlich zu verstehenden Äußerung, aber nur am Rande wahrgenommen. Die Suche ist ein Aussortieren der nicht pertinenten Belege nach dem Merkmal +- Interjektion = +- Deakzentuierung von „Gottes“ – bei gleichzeitiger Überprüfung des jeweiligen Kontextes.
Diese Suche kann auch mit Google durchgeführt werden, wo die *Snippets*, wie so oft, ihre Aussagekraft unter Beweis stellen. Es ist mit ihrer Hilfe tatsächlich möglich, die Belege nach diesen zwei Gruppen eindeutig zu ordnen.
Bei dieser Suche entsteht noch eine Restgruppe, das sind die Belege, die in keine der beiden Gruppen passen. Für den Suchenden besteht keine Veranlassung, sich näher mit diesem „Rest“ zu beschäftigen, obwohl die Mehrheit der Belege dieser dritten Gruppe angehört. Widmet man sich hingegen diesen Beispielen, dann zeigen die Korpusbelege sehr schön die Auflösung der Opposition zwischen den beiden gerade erkannten Typen, die sich dem ordnenden Blick verdanken, zugunsten einer größeren Vielfalt von Gebrauchsformen und Typen. Auffallend sind die Übergänge zwischen den Mustern, sowie das Blending der Interjektionen „In Gottes Namen“ und „Um Himmels willen“:

Alle drei waren bedeutende, international respektierte Gelehrte. Was **in Gottes Namen** konnte sie veranlassen, Hitler öffentlich, begeistert und mit wenig Zurückhaltung zu unterstützen?
Viele sowjetische Soldaten waren, wenn sie auf deutsche Zivilisten trafen, von unbändiger Wut erfüllt, die teils ihren persönlichen Leidensgeschichten entsprang, teils der Brutalität des totalen Krieges gegen die Sowjetunion. Diese Wut wuchs, je tiefer die Rote Armee nach Deutschland vorstieß, das selbst in der letzten Kriegsphase noch reich erschien im Kontrast zu den ärmlichen Verhältnissen, aus denen viele Rotarmisten kamen: »Warum **in Gottes Namen** haben sie uns überfallen? Wir besaßen doch überhaupt nichts, was sie uns hätten nehmen können!«
2008/44/Anonyma_dwdsddc.xml

So auch bei Thomas Mann, mit deutlich ironischem Unterton:

Die sieben Tage einer Woche würden dazu nicht reichen, und auch sieben Monate nicht. „Es werden, **in Gottes Namen**, ja nicht geradezu sieben Jahre sein".

Damit ist ein drittes konsistentes Muster nachgewiesen, und zwar eine zweite, von der konzessiven deutlich zu unterscheidende Interjektion. Bei diesem Gebrauch als Interjektion erhält GOTtes den Hauptton und die Interjektion ist gleichbedeutend mit „um Himmels willen". Die konzessive Bedeutung der teilweise homonymen (d. h. mit unterschiedlicher prosodischer Gestaltung) Interjektion fehlt völlig.
Widerspricht dieser Befund der Ausgangsfrage, der Frage, ob Korpusrecherchen im Sprachunterricht eine Rolle spielen können? Ich glaube, diese Frage kann man mit nein beantworten. Der unerwartet hohe Grad an Komplexität der Beleglage, kann zwar dazu verführen, die Suche an dieser Stelle abzubrechen, er könnte aber auch zum Anlass werden, die Analyse der Interjektion zu vertiefen.
Abgesehen von der literarischen Verwendung bei Enzensberger, die imstande ist, ein verschüttetes Erbe freizulegen und das Gefundene wie in einem Mausoleum der authentischen Sprache zu bewahren, ist die Erkenntnis der Übergänge, der „innovativen" Gebrauchsmuster ein sehr effizienter Zugang zur Erkenntnis, wie unsere Sprache tatsächlich funktioniert. Selbst die vorgefertigten Ausdrücke werden nicht als reine Fertigteile überliefert, sie stehen nicht isoliert in der Welt, sie sind eng eingebettet in ein Geflecht von Kontexten und Gemeintem, von stilistischen und funktionalen *Constraints*, bis hin zu idiosynkratischen Momenten des sprachlichen Handelns.
Und sie sind Veränderungen unterworfen. Diese Realität des Sprechens zu erkennen, und die dazu gehörigen sprachlichen Mittel, die optimiert sind für genau diese Art des Sprachgebrauchs, ist die beste Schulung, die ich mir für meine eingeschüchterten Studierenden des Deutschen wünschen kann. Habt keine Angst, der neuen Sprache zu begegnen, ist die Botschaft dieser Korpusbelege, in sie einzutauchen, euch von der Sprache überfluten zu lassen, alles fließt, es ist aber auch mit starken Wurzeln verankert. Diese Verankerung wahrzunehmen, ist ein entscheidender Moment im Lernprozess des Neuen. Die sprachlichen Mittel werden von Menschen gebraucht, die mit der Sprache etwas „anfangen", die mit Hilfe der Sprache handeln.
Und diese Formen, die Gebrauchsformen, verändern sich. Die Überlieferung der Interjektion „in Gottes Namen" zeigt die großen Schwankungen im Gebrauch, quantitativ und qualitativ, bis hin zur Koexistenz von zwei homony-

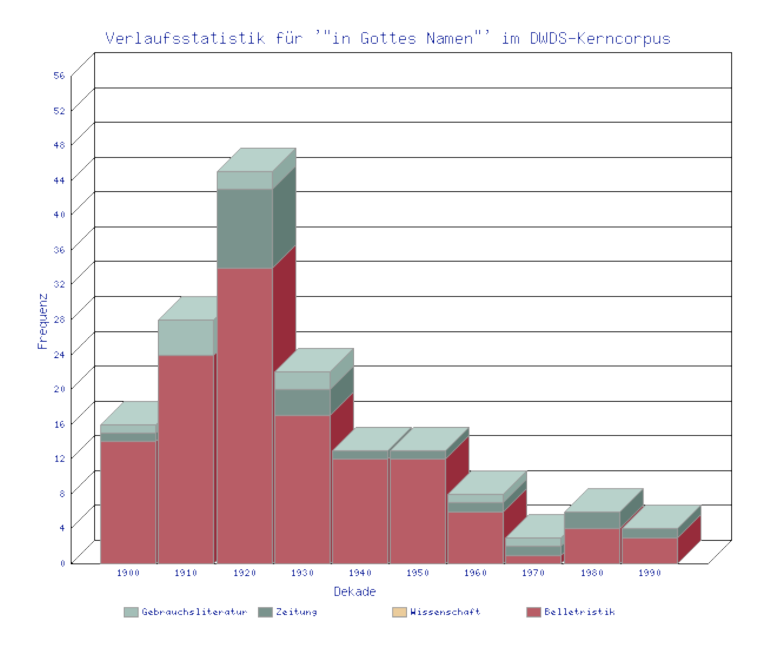
men Konjunktionen mit unterschiedlicher prosodischer Gestaltung, unterschiedlicher Bedeutung aber auch mit divergierenden Konnotationen des Gebrauchs, eine im persönlichen, privaten Bereich, im Gespräch mit vertrauten Menschen, die andere in literarisch markierter, oft auch öffentlicher Rede belegt.
Besonders auffällig wird die Interjektion schließlich beim Übersetzen. In der neuen Suchmaschine *linguee.de* wird als Übersetzung für „In Gottes Namen" der englische Ausdruck „For goodness sake" angegeben. Wenn man dann aber diesen englischen Ausdruck als Suchobjekt eingibt, erhält man kein einziges Ergebnis mit „In Gottes Namen"! Die Suche mit parallelen Texten verlangt also immer den Eingriff des bewusst agierenden und kundigen Lesers. Der Leser lernt beim Analysieren der Ergebnisse. Unter den Ergebnissen der Suche nach „In Gottes Namen" finden sich auf der einzigen Seite aus EU-Dokumenten zwei schöne Belege der konzessiven Interjektion, dazu Belege für die Bedeutung „Um Himmels willen":

Wenn es wirklich ein Problem mit dem Gebäude gäbe, dann müssten wir halt in Gottes Namen auf die andere Seite des Flusses, ins Palais de l'Europe, ziehen.
Dann ändern wir **in Gottes Namen** eben Novelfood, dann lassen wir die drei sachlichen Kriterien fallen, und immer dann, wenn irgendwo in der Kette ein genetisch veränderter Bestandteil auftritt, wird etikettiert!

Aber:

Was **in Gottes Namen** hat das alles denn mit dem Lissabon-Prozess zu tun?

Vielleicht ist es diese Parallelität, die schließlich dazu geführt hat, dass die Belege mit der Zeit an Zahl drastisch abnehmen. Die Erinnerung an typische Gebrauchssituationen verblasst, der Ausdruck wird vergessen.




Auch die Lernenden der Zweitsprache haben an diesen Prozessen der Veränderung, des Vergessens Anteil. Sie treten an die Schwelle der fremden Sprachgemeinschaft und begegnen der neuen Sprache nicht als einem unveränderbarem monolithischen Block, sondern als Korpus von Äußerungen und Sprechakten *in fieri*, zu dem sie bald selbst etwas beitragen werden.

## Literaturverzeichnis

Abutalebi, Jubin / Brambati, Simona M. / Annoni, Jean-Marie / Moro, Andrea / Cappa, Stefano F. / Perani, Daniela (Hgg.) (2007): *Switches: An Event-Related Functional Magnetic Resonance Imaging Study in Bilinguals*. In: The Journal of Neuroscience, 12.12.2007, 27, 50, 13762–13769.

Aston, Guy (Hg.) (2001): *Learning with corpora*. Bologna: Cooperativa Libraria Universitaria Editrice.

Bates, Elizabeth (1994): *Building bridges across disciplines: cognition to gene.* Address presented to the Sixth International Professional Conference of the Williams Syndrome Association, San Diego.

Bates, Elizabeth / Carnevale, George F.(1993): *New Directions in Research on Language Development*. In: Developmental Review 13, 436-470.

Bates, Elizabeth / Dale, Philip S. / Thal, Donna (1995): *Individual Differences and their Implications for Theories of Language Development.* In: Fletcher, Paul / MacWhinney, Brian (Hgg.): Handbook of Child Language. Oxford: Basil Blackwell.

Bates, Elizabeth / Goodman, Judith C. (1997): *On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-Time Processing*. In: Language and Cognitive Processes, 12 (5/6), 507-584.

Bates, Elizabeth / Elman, Jeffrey (1996): *Learning Rediscovered*. In: Science, vol. 274, 13. Dezember 1996.

Belica, Cyril (2001): *Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. http://corpora.ids-mannheim.de/ccdb/.

Berens, Heike (2009): *Usage-based and emergentist apporaches to language acquisition*. In: Linguistics 47–2 (2009), 383–411.

Bernardini, Silvia (2000): *Competence, capacity, corpora*. Bologna, Cooperativa Libraria Universitaria Editrice.

Bernardini, Silvia / Zanettin, Federico (Hgg.) (2000): *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna: CLUEB.

Bichsel, Peter (1985): *Erfahrungen beim Fremdsprachenlernen*. In: Schulmeistereien. Frankfurt a. M.: Suhrkamp, 47-59.

Bolla, Elisabetta / Drumbl, Hans (2009): *Sprachen, Kulturen, Grenzen. Teilhabe an der Sprachgemeinschaft: ein Schlüssel zur Zweitsprache.* Meran: Alpha&Beta.

Bolinger, Dwight (1976*): Meaning and Memory*. In: Forum Linguisticum. 1. 1-14.

Bubenhofer, Noah (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. In: Sprache und Wissen 4, Berlin-New York: De Gruyter.

Bühler, Karl (1934): *Sprachtheorie. Die Darstellungsfunktion der Sprache.* Stuttgart/New York: Fischer (1982).

Burckhardt, Jacob (1883): *Weltgeschichtliche Betrachtungen*, Stuttgart: Kröner, o. J.

Damasio, Antonio R. (1994): *Descartes' Error. Emotion, Reason and the Human Brain*. New York: Putnam.

Deacon, Terrence (1997): *The Symbolic Species. The Co-evolution of Language and the Human Brain*. London: Penguin.

Doughty, Catherine / Williams, Jessica (Hgg.) (1998): *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.

Drumbl, Johann (1977): *Didattica delle lingue straniere: novità e problemi*. In: Berruto, Gaetano (Hg.): Scienze del linguaggio ed educazione linguistica. Torino: Stampatori, 219-236.

Drumbl, Johann (1990): *Educazione linguistica*. In: Laeng, Mario (Hg.): Enciclopedia Pedagogica, IV. Brescia: Editrice La Scuola, 7001-5.

Drumbl, Johann (1993): *L'insegnamento delle lingue. Che cosa possiamo imparare dai metodi alternativi?*. In: L'Analisi linguistica e letteraria, I, 195-204.

Drumbl, Johann (2001): *Ein konstruktivistisches Lernmodell für Deutsch als Fremdsprache. Erfahrungen in Italien 1975-2000*. In: Meixner, Johanna / Müller, Klaus (Hgg.): Konstruktivistische Schulpraxis. Beispiele für den Unterricht. Neuwied-Berlin: Luchterhand, 125-144.

Drumbl, Johann (2002): *Aber das Eigene muss so gut gelernt seyn wie das Fremde.* In: Portmann-Tselikas, Paul R. / Schmölzer-Eibinger, Sabine (Hgg.): Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache. Textkompetenz. Neue Perspektiven für das Lernen und Lehren. Innsbruck: StudienVerlag, 247-260.

Drumbl, Johann (2003a): *Das Sprachen-Portal. Inferenz und Spracherwerb in mehrsprachiger Lernumgebung.* Meran: Alpha&Beta.

Drumbl, Johann (2003b): *Traduzione e scrittura*. Milano: LED 2003.

Drumbl, Johann (2007a): *Sprachunterricht und Ausbildung zur Mehrsprachigkeit. Bilder, Gedanken, Erinnerungen*. In: Schmölzer-Eibinger, Sabine / Weidacher, Georg (Hgg.): Textkompetenz. Eine Schlüsselkompetenz und ihre Vermittlung. Tübingen: Narr, 299-310.

Drumbl, Johann (2007b): *Intertextuelle Streifzüge im Umkreis des Übersetzens.* In: Redder, Angelika (Hg.): Diskurse und Texte. Festschrift für Konrad Ehlich zum 65. Geburtstag. Tübingen: Stauffenburg Verlag Brigitte Narr, 225-235.

Drumbl, Johann (2008a): *Leben mit Sprachen*. In: ff Südtiroler Wochenmagazin Nr. 04, 38-39.

Drumbl, Hans (2008b): *Kognition und Lernen: Neue Erkenntnisse zur Wortschatzarbeit*. In: Krumm, Hans-Jürgen / Portmann-Tselikas Paul R. (Hgg.): Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache 11:Innsbruck: Studienverlag, 89-106.

Drumbl, Johann / Federica Missaglia (1997): *Prosodie und Inferenz in zweisprachiger Lernumgebung. Neue Ansätze für Deutsch als Fremdsprache in Italien*. In: L'analisi linguistica e letteraria, V, 391-418.

Duffner, Rudolf / Näf, Anton (2006): *Digitale Textdatenbanken im Vergleich*. http://www.linguistik-online.de/28_06/duffnerNaef.html.

Edmondson, Willis / House, Juliane (1997): *Language Awareness. Fremdsprachen Lehren und Lernen.* Tübingen: Narr.

Ehlich, Konrad (2007): *Sprache und sprachliches Handeln*. Bd.1. Pragmatik und Sprachtheorie. Berlin-New York: de Gruyter.

Ellis, Nick C. (1994): *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, Nick C. (1996): *Sequencing in SLA. Phonological memory, chunking, and points of order*. In: Studies in Second Language Acquisition 18, 91-126.

Ellis, Nick C. (2001): *Memory for language*. In: Robinson, Peter (Hg.): Cognition and second language instruction. Cambridge: Cambridge University Press, 33-68.

Ellis, Nick C. (2003): *Constructions, Chunking and Connectionism. The Emergence of Second Language Structure*. In: Doughty, Catherine J. / Long, Michael H. (Hgg.): The Handbook of Second Language Acquisition. Oxford: Blackwell, 63-103.

Ellis, Nick C. (2007): *The Associative-Cognitive CREED*. In: VanPatten, Bill / Williams, Jessica (Hgg.): Theories of second language acquisition: an introduction. Mahwah, NJ: Lawrence Erlbaum, 77-95.

Ellis, Rod / Tanaka, Yoshihiro / Yamazaki, Atsuko (Hgg.) (1994): *Classroom interaction, comprehension and the acquisition of L2 word meanings*. In: Language Learning 44, 449-491.

Elsen, Hilke (1999): *Ansätze zu einer funktionalistisch-kognitiven Grammatik. Konsequenzen aus Regularitäten des Erstspracherwerbs.* Tübingen: Niemeyer.

Enzensberger, Hans Magnus (1978): *Der Untergang der Titanic. Eine Komödie*. Frankfurt am Main: Suhrkamp.

Enzensberger, Hans Magnus (1989): *The Sinking of the Titanic. A Poem.* Translated by the author, London: Paladin Grafton Books.

Evert, Stefan (2008): *Corpora and Collocations*. In: Lüdeling, Anke / Kytö, Merja (Hgg.): Corpus Linguistics. An International Handbook. Mouton de Gruyter, Berlin. vol. II, 1212-1248.

Feilke, Helmuth (1994): *Common sense-Kompetenz. Überlegungen zu einer Theorie des 'sympathischen' und 'natürlichen' Meinens und Verstehens*. Frankfurt am Main: Suhrkamp.

Feilke, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung sprachlicher Typik*. Frankfurt am Main: Suhrkamp.

Feilke, Helmuth (2007): *Syntaktische Aspekte der Phraseologie III: Construction Grammar und verwandte Ansätze.* In: H. Burger (Hg.): Phraseologie: ein internationales Handbuch zeitgenössischer Forschung, Berlin, New York: de Gruyter.

Fellbaum, Christiane (Hg.) (2007): *Idioms and Collocations. Corpus-based Linguistica and Lexicographic Studies*. New York: Continuum.

Fillmore, Charles J. (1979): *On fluency*. In: Fillmore, Charles J. / Kempler, Daniel / Wang, William (Hgg.): Individual differences in language ability and language behavior. New York: Academic Press, 85-101.

Fraas, Claudia (2001): *Usuelle Wortverbindungen als sprachliche Manifestation von Bedeutungswissen. Theoretische Begründung, methodischer Ansatz und empirische Befunde*. In: Nikula, Henrik / Drescher, Robert (Hg.): Lexikon und Text. Vaasa, 41-66. http://www.medkom.tu-chemnitz.de/mk/fraas/bedeutungswissen.pdf

Gass, Susan (1997): *Input, interaction and the second language learner*. Mahwah, N.J. Lawrence Erlbaum.

Geyken, Alexander (2004): „Korpora als Korrektiv für einsprachige Wörterbücher. In: Lili. Zeitschrift für Literaturwissenschaft und Linguistik, 136, 72-100.

Ghadessy, Mohsen / Alex Henry / Robert L. Roseberry (Hgg. 2001): *Small Corpus Studies and ELT. Theory and Practice*. Amsterdam: John Benjamins.

Grice, Paul (1989): *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.

Goldberg, Adele E. (2003): *Constructions: a new theoretical approach to language*. In: Trends in Cognitive Sciences Vol. 7,No. 5, May 2003, 219-224.

Goldberg, Adele (2006): *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gries, Stefan Th. / Wulff, Stefanie (2005): *Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora.* Annual Review of Cognitive Linguistics 3, 182-200.

Gries, Stefan Th. / Wulff, Stefanie (2009): *Psycholinguistic and corpus linguistic evidence for L2 constructions.* Annual Review of Cognitive Linguistics 7, 163-86.

Hakuta, Kenji (1974): *Prefabricated patterns and the emergence of structure in second language acquisition*. In: Language Learning, 24, 287-297.

Han, ZhaoHong (2004a): *Fossilization: Five central issues*. International Journal of Applied Linguistics, 14(2): 212-42.

Han, ZhaoHong (2004b): *Fossilization in adult second language acquisition*. Clevedon, UK: Multilingual Matters.

Han, ZhaoHong / Odlin, Terence (Hgg.) (2006): *Studies of Fossilization in Second Language Acquisition*. Clevedon: Multilingual Matters.

Han, ZhaoHong (2009): *Interlanguage and Fossilization: Towards an Analytic Model*. In: V. Cook / L. Wei (Hgg.): Contemporary applied linguistics, Vol. I: Language Teaching and Learning. London: Continuum, 137-162.

Handwerker, Brigitte / Madlener, Karin (2009): *Chunks für Deutsch als Fremdsprache. Theoretischer Hintergrund und Prototyp einer multimedialen Lernumgebung*. Hohengehren: Schneider.

Heringer, Hans Jürgen (1999): *Das höchste der Gefühle. Empirische Studien zur distributiven Semantik*. Tübingen: Stauffenburg.

Hernandez, Arturo / Ping Li / MacWhinney, Brian (2005): *The emergence of competing modules in bilingualism*. In: Trends in Cognitive Sciences Vol. 9 No. 5 May 2005, 220-225.

Hoey, Michael (2005): *Lexical Priming: A New Theory of Words and Language*, London: Routledge.

Hornung, Antonie (2002): *Zur eigenen Sprache finden*. Tübingen: Niemeyer.

Johnson-Laird, Philip N. (1983). *Mental Models. Towards a Cognitive Science of Language, Inference, and consciousness.* Cambridge, Mass.: Harvard University Press.

Kallmeyer, Werner / Zifonun, Gisela (Hgg. 2007): *Sprachkorpora – Datenmengen und Erkenntnisfortschritt* (= Institut für deutsche Sprache. Jahrbuch 2006), Berlin-New York: De Gruyter.

Kilgarriff, Adam / RychlY, Pavel / Smrz, Pavel / Tugwell, David (2004): The Sketch Engine. *http://www.kilgarriff.co.uk/Publications/2004-KilgRychlySmrzTugwell-SkEEuralex.rtf.*

Klein, Wolfgang (2000): *Prozesse des Zweitspracherwerbs.* In: Grimm, Hannelore (Hg): Enzyklopädie der Psychologie, III, Sprachentwicklung. Göttingen: Hogrefe.

Klein, Wolfgang (2001): *Typen und Konzepte des Spracherwerbs*. In: Götze, Lutz / Helbig, Gerhard / Henrici, Gert / Krumm, Hans-Jürgen (Hgg.): Deutsch als Fremdsprache, 1. Berlin-New York: de Gruyter, 604-616.

Krashen, Stephen / Scarcella, Robin C. (1978): *On routines and patterns in language acquisition and performance.* Language Learning 28.2, 283-300.

Leont'ev, Alexei A. (1971a): *Sprache, Sprechen, Sprechtätigkeit.* Stuttgart: Kohlhammer.

Leont'ev, Alexei A. (1971b): *Einige psycholinguistische Aspekte der Anfangsetappe bei der Spracherlernung*. In: Deutsch als Fremdsprache 3, 151-155.

Lieberman, Philipp (1991): *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior.* Cambridge, MA: Harvard University Press.

Lieberman, Philipp (2000): *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax and Thought*. Cambridge, MA: Harvard University Press.

Ludewig, Petra (2005): *Korpusbasiertes Kollokationslernen. Computer-Assisted Language Learning als prototypisches Anwendungsszenario der Computerlinguistik* (= Sprache, Sprechen und Computer 9), Frankfurt am Main: Lang.

Long, Michael (1991): *Focus on Form: A Design Feature in Language Teaching Methodology*. In: de Bot, Kees / Ginsberg, Ralph B. / Kramsch, Claire (Hgg): Foreign Language Research in Cross-Cultural Perspective. Amsterdam / Philadelphia: Benjamins, 39-52.

Long, Michael (2000): *Focus on Form in Task-Based Language Teaching*. In: Lambert, Richard D. / Shohamy, Elana (Hgg.): Language Policy and Pedagogy: Essays in Honor of A. Ronald Walton. Amsterdam-Philadelphia: Benjamins, 179-192.

Lüdeling, Anke / Merja Kytö (2008): *Corpus Linguistics. An International Handbook.* (= Handbücher zur Kommunikationswissenschaft 29. 1 und 2) Berlin: De Gruyter.

MacWhinney, Brian (1987): *The competition model*. In: MacWhinney, Brian (Hg.): Mechanisms of Language Acquisition. Hillsdale, NJ: Erlbaum, 249-308.

McCarthy, Michael (2004): *Touchstone*. *From Corpus to Course Book*. Cambridge: Cambridge University Press.

Meixner, Johanna / Müller, Klaus (Hgg. 2001). *Konstruktivistische Schulpraxis. Beispiele für den Unterricht.* Neuwied: Luchterhand.

Missaglia, F. (2008): *Simulierte Emotionen im Phonetikunterricht.* In: Hornung Antonie / Robustelli Cecilia (Hg): *Vivere l'interculturalita – gelebte Interkulturalität.* Tübingen: Stauffenberg, 213-223.

Meunier, Fanny / Sylviane Granger (Hgg.) (2008): *Phraseology in foreign language learning and teaching*. Amsterdam-Philadelphia: John Benjamins.

Müller, Stefan (2006): *Qualitative Korpusanalyse für die Grammatiktheorie: Introspektion vs. Korpora*. In: Zifonun, Gisela / Kallmeyer, Werner (Hgg.): IDS-Jahrbuch 2006.

Nattinger, James R. / DeCarrico, Jeanette S. (1992): *Lexical phrases and language teaching*. Oxford: Oxford University Press.

O'Keefe Anne / Michael McCarthy / Carter Ronald (2007): *From Corpus to Classroom: language use and language teaching.* Cambridge: Cambridge University Press.

Paradis, Michel (1994): *Neurolinguistic Aspects of Implicit and Explicit Memory: Implications for Bilingualism and SLA*. In: Ellis, Nick (Hg.): Implicit and Explict Learning of Languages. London: Academic Press, 393-419.

Paradis, Michel (1997): *The Cognitive Neuropsychology of Bilingualism*. In: De Groot, Anette /Kroll, Judith (Hgg.): Tutorials in Bilingualism. Psycholinguistic Perspectives. Mahwah, NJ: Erlbaum, 331-354.

Pawley, Andrew / Syder, Frances (1983): *Two puzzles for linguistic theory: nativelike selection and nativelike fluency*. In: Richards, Jack / Schmidt, Richard (Hgg.): Language and Communication. London: Longman, 191-227.

Pawley, Andrew / Syder, Frances (2000): *The one-clause-at-a-time hypothesis*. In: Riggenbach, Heidi (Hg.): Perspectives on fluency. Ann Arbor: University of Michigan Press.

Peters, Ann M. (1977): *Language learning strategies: does the whole equal the sum of the parts?* In: Language 53, 560-573.

Peters, Ann M. (1983): *The units of language acquisition*. New York: Oxford University Press.

Raupach, Manfred (1984): *Formulae in second language speech production*. In: Dechert, Hans-Wilhelm / Möhle, Dorothea / Raupach, Manfred (Hgg.): Second language productions. Tübingen: Narr, 114-137.

Pinker, Steven/Jackendoff, Ray (2005): *What's special about the human language faculty?* In: Cognition, 95(2), 201-236.

Pinker, Steven / Jackendoff, Ray (2005): *The nature of the language faculty and its implications for evolution of language* (Reply to Fitch, Hauser & Chomsky). In: Cognition*,* 97(2), 211-225.

Portmann-Tselikas, Paul R. (1991): *Schreiben und Lernen. Grundlagen der fremdsprachlichen Schreibdidaktik*. Tübingen: Niemeyer.

Portmann-Tselikas, Paul R./ Schmölzer-Eibinger, Sabine (Hgg.) (2001): *Grammatik und Sprachaufmerksamkeit.* Innsbruck: StudienVerlag.

Reisigl, Martin (1999) : *Sekundäre Interjektionen: eine diskursanalytische Annäherung.* (= Arbeiten zur Sprachanalyse, 33), Frankfurt a. M.-Berlin-Bern-New York-Paris-Wien: Peter Lang.

Riemer, Claudia (Hg.) (2000): *Kognitive Aspekte des Lehrens und Lernens von Fremdsprachen.* Festschrift für Willis J. Edmondson. Tübingen: Narr.

Rizzolatti, Giacomo / Fabbri-Destru, Maddalena / Cattaneo, Luigi (2009): *Mirror Neurons and Their Clinical Relevance*. In: Nature Clinical Practice NEUROLOGY*,* 5, 1, January 2009, 24-32.

Robinson, Peter / Roger, Gilbert (2007): *Task complexity, the Cognition Hypothesis and second language learning and performance.* International Review of Applied Linguistics 45.3, 161-76.

Römer, Ute (2008): *Corpora and language teaching*. In: Lüdeling, Anke / Kytö, Merja (2008), vol. 1, 112-131.

Schachter, Julius (1988): *Second language acquisition and universal grammar*. Applied Linguistics, 9 (3), 219-235.

Schachter Julius (1996): *Learning and triggering in adult L2 acquisition*. In Brown, Gillian / Malmkjaer, Kirsten / Williams, Jessica (Hgg.): Performance and Competence in Second Language Acquisition. Cambridge, Cambridge University Press.

Schmidt, Richard W. (1983). *Interaction, acculturation, and the acquisition of communicative competence: a case study of an adult*. In: Wolfson, Nessa / Judd, Elliott (Hgg.): Sociolinguistics and language acquisition. Rowley, MA: Newbury House, 137-74.

Schmidt, Richard W. (1990): *The role of consciousness in second language learning*. Applied Linguistics 11,129-58.

Schmitt, Norbert (Hg.). (2004): *Formulaic sequences*. Amsterdam, Philadelphia: John Benjamins.

Schmölzer-Eibinger, Sabine (2008): *Lernen in der Zweitsprache: Grundlagen und Verfahren der Förderung von Textkompetenz in mehrsprachigen Klassen***.** Tübingen: Narr.

Schmölzer-Eibinger, Sabine / Weidacher, Georg (Hgg.) (2007): *Textkompetenz: Eine Schlüsselkompetenz und ihre Vermittlung.* Tübingen: Narr.
Schumann, John H. (1994): *Emotion and Cognition in Second Language Acquisition*. In: Studies in Second Language Acquisition, 16, 2, 231-242.
Schumann, John H. (1997): *The Neurobiology of Affect in Language*. Boston: Blackwell. Zugleich in Language Learning, 48 (Supplement).
Sendlmeier, Walter F. (1989): *Aufmerksamkeitssteuerung als Methode eines Hörtrainings im Fremdsprachenunterricht*. In: Deutsche Sprache, 17, 40-51.
Selinker, Larry (1969): *Language Transfer*. In: General Linguistics, 9, 2, 671-692.
Selinker, Larry (1972): *Interlanguage*. In: IRAL. International Review of Applied Linguistics, 10, 209-31.
Selinker, Larry (1992): *Rediscovering Interlanguage*. London: Longman.
Selinker, Larry / Douglas, Dan (1985): *Wrestling with Context in Interlanguage Theory*. In: Applied Linguistics, 6, 2, 190-204.
Sinclair, John (1997): *Corpus, concordance, collocation* (= Describing English Language XVIII), Oxford: Oxford University Press.
Sinclair, John (2004): *How to use corpora in language teaching*. Amsterdam: John Benjamins.
Skehan, Peter (2009): *Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis.* Applied Linguistics 30.4, 510-32.
Spöttl, Carol / McCarthy, Michael (2004): *Comparing knowledge of formulaic sequences across L1, L2, L3, and L4.* In: Schmitt, Norbert (Hg.): Formulaic sequences. Amsterdam-Philadelphia: John Benjamins, 191-225.
Sperber, Dan / Wilson, Deirdre (1986, 1995): *Relevance*. Oxford: Blackwell.
Steyer, Kathrin (2004): *Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven.* In: Kathrin Steyer (Hrsg.): Wortverbindungen - mehr oder weniger fest (= Jahrbuch des Instituts für Deutsche Sprache 2003), Berlin-New York: de Gruyter, 87-116.
Tomasello, Michael (2006): *Construction Grammar for Kids*. In: Constructions, 1, 11 / 2006.
Tomasello, Michael (2009): *Die Ursprünge der menschlichen Kommunikation*. Frankfurt am Main: Suhrkamp. (*Origins of Human Communication*. MIT Press: 2008).
Tomasello, Michael (2003): *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.
Tomasello, Michael / Bates, Elizabeth (Hgg.) (2001): *Language Development: the essential readings*. Oxford, Blackwell.
Valéry, Paul (1992): *Cahiers/Hefte*. Bd. 5. Dt. Übers. v. R. Huschke, H. Köhler und J. Schmidt-Radefeldt. Frankfurt a. M: Fischer.

Van Lier, L. (2000): *From Input to Affordance: Social Interactive Learning from an Ecological Perspective*. In: Lantolf, James P. (Hg.): Sociocultural Theory and Second Language Learning: Recent Advances. Oxford: Oxford University Press, 245-259.

Van Lier, L. (2004): *The Ecology and Semiotics of Language Learning: A Sociocultural Perspective.* Boston: Kluwer Academic Publishers.

VanPatten, Bill (2002): *Processing Instruction: An Update*. In: Language Learning, 52, 755-803.

VanPatten, Bill / Williams, Jessica / Rott, Susanne / Overstreet, Mark (Hgg.) (2004): *Form-Meaning Connections in Second Language Acquisition.* Mahwah, N.Y.: Lawrence Erlbaum.

Vihman, Marilyn (1982): *Formulas in first and second language acquisition.* In: Obler, Loraine K. / Menn, Lise (Hg.): Exceptional language and linguistics. New York: Academic Press, 261-284.

Weinert, Regina (1995): *The role of formulaic language in second language acquisition: a review*. In: Applied Linguistics, 16 (2), 180-205.

Weinert, Sabine (1991): *Spracherwerb und implizites Lernen. Studien zum Erwerb sprachanaloger Regeln bei Erwachsenen, sprachunauffälligen und dysphasisch-sprachgestörten Kindern.* Bern-Göttingen-Toronto: Huber.

Weinert, Sabine (2009): *Implicit and explicit modes of learning: similarities and differences from a developmental perspective*. In: Linguistics 47–2 (2009), 241–271.

Wong Fillmore, Lily (1976): *The second time around. Cognitive and social strategies in second language acquisition*. Stanford: Stanford University.

Vygotskij, Lev S. (2002): *Denken und Sprechen. Psychologische Untersuchungen*. Hg. und aus dem Russischen übersetzt von Joachim Lompscher und Georg Rückriem. Mit einem Nachwort von Alexandre Métraux. Weinheim, Basel: Beltz.

Weinreich, Uriel (1953): *Languages in Contact: Findings and Problems.* New York: Publications of the Linguistic Circle of New York, 1. (Lingue in contatto. Con saggi di Francescato / Grassi / Heilmann. Torino: Boringhieri, 1974. Sprachen in Kontakt: Ergebnisse und Probleme der Zweisprachigkeitsforschung. Mit einem Vorwort von A. Martinet. München: Beck, 1976.)

Wray, Alison (2000): *Formulaic Sequences in Second Language Teaching: Principle and Practice.* Applied Linguistics 21/4, 463-489.

# Korpusinstrumente im Umkreis des Lernens

**Renata Zanin – Freie Universität Bozen**

## Abstract

Neuere Beiträge zur Wortschatzvermittlung im gesteuerten DaF-Unterricht gehen über die Wortgrenze hinaus und setzen als Grundelement der zu erlernenden Sprache *Konstruktionen* (Goldberg) und den *Ausdruck* als *Kristallisationspunkt für Kombinations- und Selektionspräferenzen* (Feilke) an. Dieser neue Blick auf die Sprache kann mit Hilfe von Korpusrecherchen in den Sprachunterricht eingebracht werden. Das umfassende Ziel, den gesteuerten Spracherwerb als Data Driven Learning mit Hilfe von Korpusrecherchen auf eine neue Grundlage zu stellen, wurde nicht erreicht und die bisher vorgelegten Erfahrungsberichte betreffen kaum mehr als Nischenangebote im Panorama des Fremdsprachenunterrichts. In diesem Beitrag werden positive didaktische Erfahrungen mit Korpora vorgestellt und analysiert, die in der dreisprachigen Arbeitsumgebung der Freien Universität Bozen entstanden sind. Gesteuerte Korpusrecherchen mit einem gemeinsamen Interface für Korpora in den drei Arbeitssprachen Deutsch, Italienisch und Englisch erweisen sich dabei nicht nur als ansprechendes Angebot für fortgeschrittene Studierende, sondern auch als Werkzeug für die Erarbeitung einer grundlegend neuen Haltung der italienischen Studierenden gegenüber dem Ziel des Erlernens der Zweitsprache Deutsch und allgemein, für den Umgang mit Sprachen in Situationen des täglichen Sprachkontakts und des Schreibens in der Zweitsprache.

## Einleitung

Das Interesse, das Lehrpersonen und Lernende den großen Sammlungen von Texten entgegenbringen, ist nur zu einem ganz geringen Teil identisch mit dem Interesse der wissenschaftlichen Erforschung dieser Sammlungen nach korpuslinguistischen Fragestellungen und mit korpuslinguistischen Methoden. Ein ausgewiesener Spezialist solcher Untersuchungen, Stefan Gries der University of California, Santa Barbara, schreibt dazu:

> By its very nature, corpus linguistics is a distributional discipline. In fact, it has been argued that corpora as such contain nothing but distributional frequency data. (Gries 2010:5)

Wenn man vom Interesse an der reinen Frequenz von Mustern, Kollokationen und Konstruktionen absieht, was für eine erkenntnisleitende Fragestellung verbleibt dann überhaupt und aus welchem Grund sollte man sich mit Korpo-

ra im Unterricht befassen? In einem Forumsbeitrag hat Adam Kilgariff, der die äußerst benutzerfreundliche SketchEngine entwickelt hat, ein resignierendes Resumee gegeben:

> Concordances are too tough for learners. So DDL has not taken off. After 20 years, it remains a tiny minority interest.

Diese Feststellung steht wie eine unbequeme Barriere vor der eigenen Arbeit mit Korpora im Sprachunterricht und fordert zu grundlegenden Überlegungen heraus. Eine Konsequenz, die aus der Komplexität von Korpusrecherchen abgeleitet wird: Korpusinstrumente gehören nur in die Hand von sehr fortgeschrittenen Lernenden, sagt man, oder von Muttersprachlern, die sich vorgenommen haben, feinste stilistische Varianten auszuloten. Aus der Gesamtmenge von distributionellen Daten werden also gleichsam Nischen-Daten herausgefiltert, die sprachliche Phänomene an der Grenze der Grammatik und Stilistik betreffen und die selbst in großen Textsammlungen, wenn man nach Textsorten und Registern differenziert, nicht in großer Zahl dokumentiert sind.
Im Sprachunterricht sind also nur Einzelfälle relevant, die nach ganz bestimmten Kriterien der Relevanz aus der Masse der Gesamtbelege herausgefiltert werden.
Das Thema dieses Aufsatzes ist die eng begrenzte Methode der Korpusanalyse in genau dieser Hinsicht. Ich berichte über didaktische Erfahrungen mit Korpora im DaF-Unterricht mit ganz bestimmten Kategorien von Lernenden, aber die Arbeit mit Korpusrecherchen wird vor den Hintergrund einer Neubestimmung didaktischer Prioritäten auf dem Gebiet des Wortschatz- und Grammatikunterrichts gestellt. Die ausgrenzende Definition von Korpusanalysen, die Gries als anerkannte Wahrheit ausspricht, hat nämlich sehr wohl eine Verbindung zum Sprachunterricht, und zwar über die Brücke des dem Unterricht zugrunde gelegten Grammatikmodells. Zu den strukturalistischen Modellen, die großes Augenmerk auf die kleinsten Einheiten legen, ist jüngst die Definition der Sprachbefähigung im engen Sinn gekommen (Hauser, Chomsky, Fitch 2002), die in der Bestimmung mündet: *In fact, we propose in this hypothesis that FLN comprises only the core computational mechanisms of recursion as they appear in narrow syntax and the mappings to the interfaces*. Aus dieser *Sicht* kann ein *Bild* der Sprachbefähigung abgeleitet werden, das den Umgang mit Sprachen bis hinein zur Bestimmung von Übungsstrategien zum gesteuerten Erwerb der Zweit- und der Fremdsprache führt.
Dieses Bild scheint heute allerdings vielen Sprachwissenschaftlern und dem Großteil der Sprachdidaktiker und der Lehrpersonen zu eng gefasst.

Die Bestimmung eines Kernstücks, *core*, sprachlicher Funktionalität und die Gleichsetzung dieses Kerns mit der Syntax in engem Sinn, führt zum Ausschluss all jener sprachlichen Phänomene aus dem Kernbereich theoretischer Fragestellungen, die in Chomskys Sicht einer unbedeutenden *Peripherie* angehören. Dazu gehören alle holistischen Phänomene der Sprache, Mehr-Wort-Ausdrücke, idiomatische Ausdrücke, Valenzstrukturen, sprachliche Konstruktionen, wie sie auch beim kindlichen Spracherwerb zu finden sind.
Diese Ausdrücke lassen sich in der Folge der wissenschaftlichen Diskussion der letzten 20 Jahre nicht mehr an die Peripherie der Sprachen und der Sprachwissenschaft verlegen. Auf dem Höhepunkt der jüngst geführten kritischen Auseinandersetzung zur Frage nach dem wesentlichen Merkmal der menschlichen Sprachbefähigung äußerten sich Pinker und Jackendoff zur Rolle von Idioms und Mehr-Wort-Konstruktionen:

> First, idioms and constructions are as specific to language as any other syntactic phenomenon. That is, they do not come for free with understanding the concepts underlying word meanings, and therefore cannot be relegated to FLB.
> Second, idioms and constructions are not "peripheral" to language on any ordinary understanding of that word. We know of no accurate estimates, but the number of idioms and constructions that speakers know appears to be of comparable magnitude to the number of words. Furthermore, an examination of any stretch of actual language shows that idioms and constructions make up a substantial proportion of speakers' usage. (Cognition 97, 2005: 221)

Dieses Plädoyer zugunsten der idiomatisch geformten Sprache, also jener sprachlichen Phänomene, die im Alltag gehäuft auftreten und im Sprachunterricht sowohl der Erst- als auch der Zweitsprache Deutsch besonders markant in Erscheinung treten, knüpft ein Band zwischen Sprachtheorie und Sprachlern-Praxis, wie es in den letzten fünfzig Jahren nicht zur Verfügung stand. Die Praxis und die Instrumente, die für die Praxis Bedeutung haben, stehen somit nicht mehr im Ruf, sich mit irrelevanten Epiphänomenen der Sprache abzugeben.
Eine solche Neugewichtung hat Folgen für das theoretische Fundament der Sprachwissenschaft und für die Sprachdidaktik, die sich nun tatsächlich durchzusetzen beginnen:

> Sollte der sprachliche Normalfall nicht das Okkasionelle, sondern das Geprägte und Vorgeprägte sein, müssten Theorien und Modelle der Linguistik darauf umgestellt werden, dass die ad-hoc-Bildungen mehr oder weniger kreative Realisierungen idiomatisch geprägter und vorgeprägter Modelle darstellen, dass also die grammatischen Regeln auf der Basis dieser Modelle funktionieren. (Ágel 2004: 67)

Von dieser Umwertung der Prioritäten und des theoretischen Blicks auf die Sprachen aus können Impulse für die Sprachdidaktik abgeleitet werden, die ich nun schrittweise nachvollziehen möchte. Als Ausgangspunkt für die Analyse der Schnittstelle zwischen traditioneller Grammatik und korpusorientierter Sprachbetrachtung bieten sich die Phraseologismen an, die Feilke operational definiert:

Idiomatische Prägung liegt vor, insofern in einer Sprache ausdrucksseitige Kombinations- und Selektionspräferenzen und die Etablierung kompositionell nicht prädiktabler, präferentieller Bedeutung von Ausdrücken zu belegen sind. (Feilke 1998: 74)

Das Wort stellt ein Verknüpfungspotential bereit, das die strukturell möglichen Alternativen ermöglicht – während die Äußerung definitionsgemäß über das Sprachliche im engen Sinn hinausreicht und die illokutionäre Sphäre zur Bestimmung von Selektivitätsmustern einbezieht (Feilke 1996: 68). Es werden dabei kognitive Bereiche angesprochen und aktiviert, die bei jeder Sprechtätigkeit eine Rolle spielen, deren Beitrag zum Glücken der Sprechakte im Rahmen des institutionellen Lernens aber bisher kaum zur Kenntnis genommen wurde.
Auf dem Gebiet der Grammatiktheorie hat eine Umkehr zu Gunsten größerer Einheiten als Grundelemente der Sprache begonnen, die in der Construction Grammar eine umfassende Definition gefunden hat.

Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg 2006:5)

Sofort einsichtig sind die Konsequenzen einer solchen Sicht auf die Sprachen für jene Teile des Syllabus, die der Stärkung der Sprachaufmerksamkeit gewidmet sind, dem Aufbau und der Absicherung des Sprachgefühls und des Sprachgebrauchs (fluency and native-like selection).
Aufmerksamkeit auf die Sprache, eine Sprache, die zu anderen Sprachen in engem Kontakt steht und den Einflüssen dieser Sprachen ausgesetzt ist. Das ist in diesem Fall das Italienische, aber die Beobachtungen und Überlegungen, die hier diskutiert werden, gelten analog für andere Situationen des Sprachkontakts. Es braucht eine Fülle von wohlüberlegten didaktischen Bemühungen, um die Aufmerksamkeit auf den Bereich der Mehr-Wort-Konstruktionen in beiden, oder in allen Schulsprachen konsequent zu fördern. Die Studierenden von Sprachen müssen Gelegenheit finden, Sprecher und normale Nutzer der zwei-

ten Sprache zu werden und sie brauchen auf dem Weg zu diesem Ziel Unterstützung durch Unterricht und gesteuerte Übungsstrategien. Der Gegenstand des Lernens und Übens ist nicht allein die neue Sprache, sondern die Übung einer Haltung zu den Sprachen und zum Spracherwerb, die in einen umfassenden kognitiven Rahmen eingebettet ist. Diese Rahmenbedingungen sind zu schaffen und die Lernenden sind darauf vorzubereiten, Wege des Übens einzuschlagen, die ihren bisherigen Lernerfahrungen in keiner Weise entsprechen. Eine solche Veränderung im Lernverhalten ist ein langwieriges Unterfangen und nur mit großem Aufwand zu erreichen.
Sprache ist ein zentrales Thema für Kultur, Bildung und Schule. Ein Thema mit einem unermesslichen Potential an Wissensbedarf, Forschung sowie kulturpolitischen Aktivitäten und Entscheidungen. Dieser Anspruch kollidiert mit der immer stärker sichtbaren Ressourcen-Knappheit. Aufmerksamkeit und Lebenszeit sind zu Zielen der Anbieter von Medieninhalten geworden. In allen Bereichen stellt die Verteilung der Ressourcen ein brisantes Thema dar. Der Kampf um die Ressource Aufmerksamkeit betrifft nicht nur die Medien und den ungeübten Nutzer eines ins Unermessliche gestiegenen Überangebots, sondern auch Forscher, Wissenschaftler, Lehrpersonen und Anwender von Theorien. Auch sie sind einem Überangebot ausgesetzt und gezwungen, zu wählen. Die Schritte in die Praxis und in der Praxis sind abhängig von den Fragen, die wir stellen, von den Büchern, die wir lesen und den Prioritäten, die wir auf der Grundlage des geteilten Wissens setzen. Mit klaren Worten: Wir können es uns nicht leisten, die Ressourcen an Zeit und an Aufmerksamkeit mit nicht adäquaten Lernobjekten und Lernstrategien zu vergeuden.
Philip Lieberman hat in seinem Buch „Human Language and Our Reptilian Brain. The Subcortical Bases of Speech, Syntax and Thought" (Harvard University Press 2000), in dem viele der heute aktuellen Themen im Bereich der Sprache und der Kognitionsforschung antizipiert werden, Schlüsse gezogen, die auf unser Handeln im Bereich des Sprachunterrichts direkte Folgen haben:

> If language were an instinct, then only minimal attention would have to be paid to the circumstances in which children are raised. In this light, spending public funds to enrich the lives of children born into poverty would be unnecessary.
> As the evidence discussed in this book demonstrates, language is a skill. Like other skills, such as playing the violin, certain neural and anatomical mechanisms are required. But like other skills, language is honed by practice in places and circumstances conducive to learning. (166)

Wie heftig umkämpft solche Äußerungen im Jahr 2000 noch waren, klingt im Umkreis der zitierten Worte an, wo Liebermann sich von der These distan-

ziert, Sprache sei ein *Instinkt* und umfassende bildungspolitische Konsequenzen aus der von ihm aufgezeigten neuen Sicht auf die Sprache und den Spracherwerb ableitet, die im lakonischen Satz münden: *The message here is that we as a nation have not done enough.*
Ich möchte für meinen Bericht über Anwendungen von Korpusinstrumenten im Sprachunterricht in der mehrsprachigen Lernsituation, wie sie für Südtirol charakteristisch ist, nun den kleinen Abschnitt aus dem Zitat nutzen, der uns in medias res der Arbeit begleiten kann: *But like other skills, language is honed by practice in places and circumstances conducive to learning.*
Der Schwerpunkt liegt bei der Frage, ob in ausreichendem Maße dafür gesorgt ist, dass die Lernumgebung, in der unsere Schüler die von ihnen erwarteten Leistungen erbringen sollen, tatsächlich produktives Lernen zu fördern imstande ist und welche Aktivitäten des Lernens und des Übens bevorzugt angestrebt werden sollen – wobei ich gleich an ein Wort erinnern möchte, das Konrad Ehlich auf dem Neustifter Kolloquium im Januar 2008 gesprochen hat: *Sprache kann nicht als Unterrichtsgegenstand gelehrt und gelernt werden*.
In der für Südtirol typischen Lernsituation können wir davon ausgehen, dass für das Erlernen der Zweitsprache überdurchschnittlich viel Zeit zur Verfügung steht. Von der Grundschule bis zum Abitur absolvieren die Schüler insgesamt nicht weniger als 2.500 Stunden Sprachunterricht. Das ist, gemessen am Stundenkontingent der öffentlichen Schulen in Italien, sehr viel. Gemessen am zeitlichen Aufwand, den das Kind benötigt, um die erste Sprache zu lernen, ist es aber immer noch unerhört wenig. Sprachenlernen im fortgeschrittenen Alter, also jenseits des kognitiven Fensters der ersten Jahre nach der Geburt, braucht besondere Unterstützung und Hilfestellungen.
Worin diese Unterstützung bestehen soll, kann mit einem Satz umschrieben werden: entscheidend für effizientes Lernen ist das Ermöglichen von sehr vielen sinnhaften Begegnungen und Erlebnissen mit der neuen Sprache. Mit dem Hinweis auf den *comprehended input* (Gasser, 1994) ist dabei so etwas wie ein unterer, minimaler Schwellenwert angesprochen. Denn zum reinen Verstehen kommt noch die emotionelle Erfahrung sinnhaften Handelns und etwas, was man altmodisch mit *Freude am Handeln und am Lernen* bezeichnen kann.
Das führt in direktem Anschluss an das neue Grammatik-Modell zu Lern- und Übungsstrategien, die Wege des Lernens fördern wie das Beobachten und Vergleichen von Erfahrungen, Analogieschlüsse, inferenzielle Schlüsse, die metasprachliche und metakommunikative *skills* verlangen, die sich im Lauf der Zeit zu brauchbaren Arbeits- und Lerninstrumenten entwickeln. Den formelhaften Ausdrücken der Sprache kommt dabei eine besondere Bedeutung zu. Sie stellen eine Brücke dar zur authentischen Sprache der Menschen, die diese

Sprache im Alltag sprechen, und die die Lernenden als die andere Sprachgruppe vor Augen haben. Die Sprachgruppe, der sie selbst nicht angehören. Die Erfahrung des Trennenden schlechthin. Erfolgreiches Erlernen der formelhaften Sprache eröffnet einen Weg zur anderen Sprachgruppe, wie neuere Forschungen nahelegen, die eine positive Korrelation zwischen dem erfolgreichen Lernen der *formulaic language* und der sozialen Integration der Lernenden festgestellt haben. (Adolphs/Durow 2004; siehe auch Dörnyei et al. 2004)

## 1. Die Aufmerksamkeit schulen

Vor diesem Hintergrund möchte ich einige Momente des Lernens mit Hilfe von Korpusinstrumenten aus unserer sprachdidaktischen Praxis an der Freien Universität Bozen referieren. Nach mehrjähriger Erprobung der Korpusinstrumente einer eigens für die Sprachdidaktik optimierten Plattform, hat sich folgende Anordnung der Arbeits- und Übungsstrategien als vorteilhaft herausgestellt:
Jede Unterrichtseinheit besteht aus 7 Schritten:

1. Schritt (Lehrer)
   Wahl eines Textes (Kriterien der Wahl: Inhalt, Sprache, Textaufbau)
   Suche nach Vergleichstexten und Vorbereitung eines Dossiers zum Thema;
2. Schritt (Klasse)
   Lektüre des Ausgangstextes in der Klasse, erste Erkundung des sprachlichen Umfeldes durch spontane Suche (Snippets in Google), Brainstorming (Was weiß ich schon zum Thema), Vergleichstexte, Erarbeitung von Ein- oder Zwei-Satz-Zusammenfassungen mit Hilfe von Bausteinen des ersten Textes;
3. Schritt (Klasse oder zu Hause):
   Zusammenfassung und Kommentar (jeweils max. 60 Wörter);
4. Schritt (Lehrer)
   Korrektur der Arbeiten, Anonymisierung der Schülerarbeiten;
5. Schritt (Lehrer und Schüler)
   Besprechung einer oder mehrerer Schülerarbeiten in der Klasse, Fokussierung sprachlicher Charakteristiken und Muster (ebenso Musterabweichungen), Strategien werden transparent gemacht, autonome Suche im Korpus wird angeregt und unterstützt (Suchanweisungen - Unterstreichungen - werden gegeben);
6. Schritt (Schüler):
   Ergebnisprotokoll der Suche, Vertiefung der Suche, Serendipity-Prinzip;
7. Schritt (Lehrer oder Schüler)
   Evaluation und/oder Peer-reviewing.

Ein typischer Text aus dem hier angesprochenen Umfeld des Lernens und des Sprachgebrauchs ist folgender Kommentar einer Südtiroler Studentin zu einem Bericht über die „Deutschpflicht“ an der Hoover-Schule in Berlin. Der Ausgangstext ist ein Interview in der ZEIT.

**Die Zeit -** Chancen Schule - Nr. 26 - 21. Juni 2007

Gefragt - Nezir Asanovic, 16, Schulsprecher an der Herbert Hoover-Gesamtschule in Berlin-Wedding

**Wir sprechen Deutsch**

***Seit anderthalb Jahren müssen an Ihrer Schule alle Schüler Deutsch sprechen, egal ob im Unterricht oder auf dem Schulhof. Hat die Deutschpflicht das Klima an der Schule verändert?***

Es ist viel friedlicher geworden an unserer Schule, die Atmosphäre hat sich gewandelt. Es gab in diesem ganzen Schuljahr keine einzige Schlägerei. Früher gab es etliche, die oft nur durch Missverständnisse entstanden sind.

***Durch sprachliche Missverständnisse?***

Ja. Wir haben einen Ausländeranteil von 90 Prozent an unserer Schule, hier mischen sich rund 16 Nationalitäten. Da rufen ein paar Türken einem Araber was auf Türkisch hinterher, der es gar nicht oder falsch versteht, und schon fühlt er sich in seinem Stolz verletzt und muss sich verteidigen. Wir waren selbst überrascht, dass viele körperliche Auseinandersetzungen so viel mit Sprache zu tun hatten. Seitdem alle Deutsch miteinander sprechen, entstehen bestimmte Konflikte gar nicht erst.

***Wirklich alle?***

Natürlich gibt es immer Schüler, die provozieren wollen. Das war ganz extrem, als die Deutschpflicht eingeführt wurde. Viele Schüler fühlten sich angegriffen, fanden die Entscheidung rassistisch. Aber ihr Protest zielte bald ins Leere. Wir haben viel mit Lehrern diskutiert, und die Mehrheit der Schüler war sehr schnell vom Sinn dieser Regel überzeugt.

***Was passiert mit denen, die trotzdem nicht Deutsch reden?***

Heute passiert das kaum noch. Das Deutsche gehört jetzt zum Schulalltag. Bestrafungen gab es von Anfang an nicht, nur Ermahnungen. Ich glaube, auch deshalb verlor das Provozieren sehr schnell seinen Reiz.

***Hatte die neue Regel auch Auswirkungen auf den Unterricht?***

Ich glaube, das Leistungsniveau in unserer Klasse ist gestiegen. Wir reden auf einem ganz anderen Niveau miteinander. Diskutieren plötzlich über Texte und Gedichte, alleine unter Schülern. So was gab’s früher kaum. Da war es vielen peinlich, etwas Falsches auf Deutsch zu sagen. Heute gewinnt man mit jedem Tag mehr Selbstbewusstsein.

**Sind sich die vielen Nationen nähergekommen?**

Das multikulturelle Gefühl an der Schule ist stärker geworden. Es ist jetzt egal, ob einer Türke ist oder Araber. Dafür steht auch, dass ich als Serbe überhaupt Schulsprecher werden konnte – trotz unserer großen türkischen Mehrheit.

**Schülerarbeit**

**Kommentar: Wir sprechen Deutsch**

Meiner Meinung nach gibt es in Bezug auf die Einführung der Deutsche Sprache einige positive Aspekte.
Ich denke, dass eine einzige Sprache viele Probleme in der Schule lösen kann.
In diesem Fall hat die Sprache verschiedene Kulturen, Nationalitäten und Einstellungen verbunden.
Dank der Sprache fühlen sie sich Teil der selben Gruppe.
Da die Schüler verpflichtet sind, ausschließlich Deutsch zu sprechen, nutzen sie die Deutsche Sprache und verbessern deren Sprachkompetenz.
Im Allgemeinen ist diese Pflicht positiv, weil die Schule immer etwas Konkretes machen sollte, um die Schwierigkeiten zu überwinden.

Die mangelnde Textualität ist, wie auch in sehr vielen anderen Fällen, schon am Zeilenumbruch nach jedem Satz zu erkennen. Jeder Satz steht wie eine Insel für sich, ohne jede Verbindung zu den anderen Sätzen oder zu einem wie immer gearteten Textzusammenhang. Die Isolation der Einzelsätze verführt dazu, die Sätze tatsächlich in ihrer Isolation wahrzunehmen und nach rein formalen Gesichtspunkten zu bewerten.
Im Rahmen unseres Lernmodells mit Korpusinstrumenten sind zwei Komponenten klar zu unterscheiden. Erstens die Schulung der Sprachaufmerksamkeit, die von Fall zu Fall in allen drei Schulsprachen erfolgt, und zweitens der Aufbau der für die *Resonanz* nötigen kognitiv-sprachlichen *Infrastruktur*, wo die Lexik in Form von verknüpften Konstruktionen und *chunks* im weitesten Sinn ihren Ort finden kann (Drumbl, in diesem Band). Für die Studierenden nach 13 Jahren Schulunterricht geht es darum, den Ort für die fehlende Grundlage einzurichten und auszustatten. Lernende in der ersten Phase des Lernens können hingegen von Anfang an damit beginnen. Das ist der Grund dafür, weshalb wir den Einsatz von Korpusinstrumenten im Anfängerunterricht nicht ausschließen.
Hier wird die Übung mit Studierenden wiedergegeben, die 13 Jahre Schulunterricht hinter sich haben und in einem zweisprachigen Studiengang in Südtirol studieren. Die Übung hat das Ziel, die Schärfung der Sprachaufmerksamkeit mit dem Aufbau der fehlenden oder extrem fragilen sprachlichen Grundlage zu verbinden. Als Anlass könnte das fehlende *n* der Adjektivendung erscheinen, das allerdings nicht einfach als Korrektur in die Arbeit eingetragen wird, sondern zum Auslöser einer gezielten Korpusrecherche wird:
[…] die Einführung der Deutsche Sprache […]
Die Übung ist Anlass, zwischen den beiden Eingabefenstern unterscheiden zu lernen, dem Fenster für die Eingabe eines Lemmas und dem Fenster für die Eingabe von Wortverbindungen oder von einzelnen Wortformen. Beim Suchen ist es von grundlegender Bedeutung, das, was der Suchende mit großer Sicherheit weiß und kann, von dem zu trennen, wo noch Unsicherheit

herrscht. Voraussetzung ist daher die Kenntnis der Lemma-Form der gesuchten Wörter. Im Zweifel kann die Form probeweise als Lemma eingegeben und auf diese Weise überprüft werden. In der ersten Phase der Arbeit mit Korpusbelegen ist die Suche mit Lemmata der Suche nach festen Mehr-Wort-Gruppen (Phrase-Fenster) vorzuziehen.

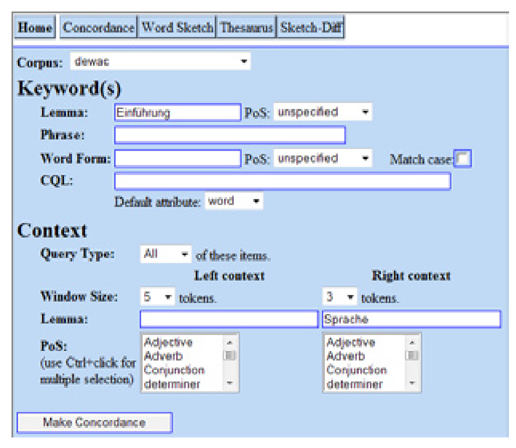



Abb. 1: dewac – Korpus Südtirol

Die erste Gruppe wird daher in die beiden Nominal-Lemmata aufgelöst und als *Einführung* + *Sprache rechts Abstand 3* eingegeben.




Abb. 2: dewac – Korpus Südtirol

Bei den Ergebnissen wird die korrekte Adjektivendung im Genitiv angezeigt, deren Korrektur aber nicht das Ziel der Aufgabe darstellt. In der Einführungsphase kann also unauffällig auf diese hochfrequente Fehlerquelle hingewiesen werden.
Sogleich kann der Abstands-Parameter transparent gemacht werden. Bei *Abstand 1* werden keine Wörter zwischen den beiden Nomina angezeigt, also auch keine Adjektive, die für die Lernenden von Interesse sind. Also muss der Abstand vergrößert werden, wie bei dieser Suche, die mit *Abstand 3* durchgeführt wurde, wobei zwei deutlich unterschiedliche Muster erkennbar werden: *Einführung der* und *Einführung in die […].*
Die computertechnische Einführung in die Bedienung der Software stellt also schon einen unauffälligen Moment des Erlernens der Sprache dar.
Nach dieser Einführung ist es nicht mehr notwendig, die Abweichungen innerhalb der Nominalgruppen immer wieder als eigene Suche vorzuschlagen. Wir formalisieren solche Vorschläge zum eigenständigen Suchen, indem wir in den Schülerarbeiten vier Stellen unterstreichen, zwei korrekte und zwei fehlerhafte. Die Lernenden müssen selbst herausfinden, welche Stellen korrekt formuliert sind. Die Überprüfung dient dabei nicht eigentlich der Fehlerkorrektur, sondern der Suche nach rekurrenten Mustern und dem Vergleich der gefundenen Muster mit dem selbst geschriebenen Satz oder der Wortgruppe.
Solche Recherchen werden nach einigem Üben selbstständig gemacht. Vor allem bei der Arbeit in Kleingruppen funktioniert das selbst-gesteuerte Suchen, jeder *Fund* motiviert zu weiteren Suchaktionen, die sodann in rudimentärer Form dokumentiert werden. Durch Unterstreichen von vier Mehr-Wort-Gruppen pro Schülertext werden Suchaufgaben aufgegeben, die individuell oder in Kleingruppen gelöst werden können. Es werden auch sprachlich unauffällige Gruppen markiert, so dass die Recherche nicht den Status einer Fehlerkorrektur erhält, sondern den einer Suche nach dem Potential von Konstruktionen mit ihren Variationen und ihren unterschiedlichen Kontexten. Die Studenten müssen entscheiden, ob sie durch Eingabe in das PHRASE-Fenster nach festen Verbindungen suchen wollen, oder nach zwei oder drei Wörtern mit Angabe der Position und des Abstandsparameters. Hier zwei Angaben im ersten Satz des Schülertextes, die unterschiedliche Suchstrategien verlangen.
Die erste Gruppe, als *Einführung + Sprache rechts Abstand 3* eingegeben, ergibt ein differenziertes Ergebnis mit relevanten Beispielen für *Amtssprache* und für *Unterrichtssprache*. Ganz nebenbei wird die Adjektivendung angezeigt, deren Korrektur aber nicht das Ziel der Aufgabe darstellt.
In bestimmten Fällen könnte diese Gruppe auch ohne Markierung bleiben, um den Eindruck zu vermeiden, das Aufsuchen von Fehlern sei das Ziel der

Arbeit. Die erste Recherche ist somit der auf den ersten Blick erkennbaren Gruppe *einige positive Aspekte* gewidmet,
Meiner Meinung nach gibt es in Bezug auf die Einführung der Deutsche Sprache einige positive Aspekte.
die als Konstruktion in das Phrase-Fenster eingetippt werden kann, falls die Studierenden die Gruppe mit großer Sicherheit als korrekt erkennen.

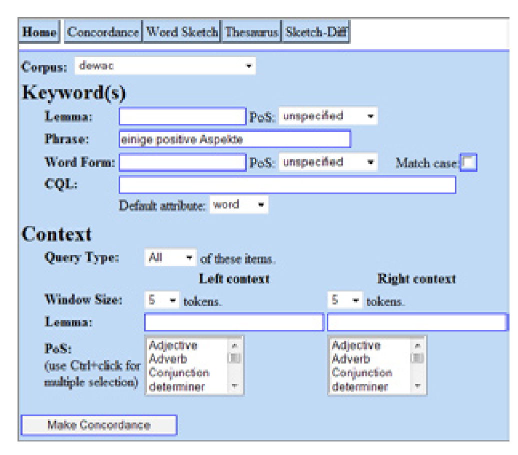



Abb. 3: dewac – Korpus Südtirol

Die Suche endet mit einer Überraschung. Mit Hilfe der Korpussuche können nämlich Gebrauchsbeschränkungen erkannt werden, die selbst versierten Sprechern des Deutschen nicht immer bewusst sind. Der chunk *einige positive Aspekte* wird in allen Belegen aus dem Korpus von einem obligatorischen negativen Hintergrund abgehoben. Der negative Hintergrund gehört demnach zum *Glücken* des Sprechaktes, in dem diese Wortverbindung vorkommt.

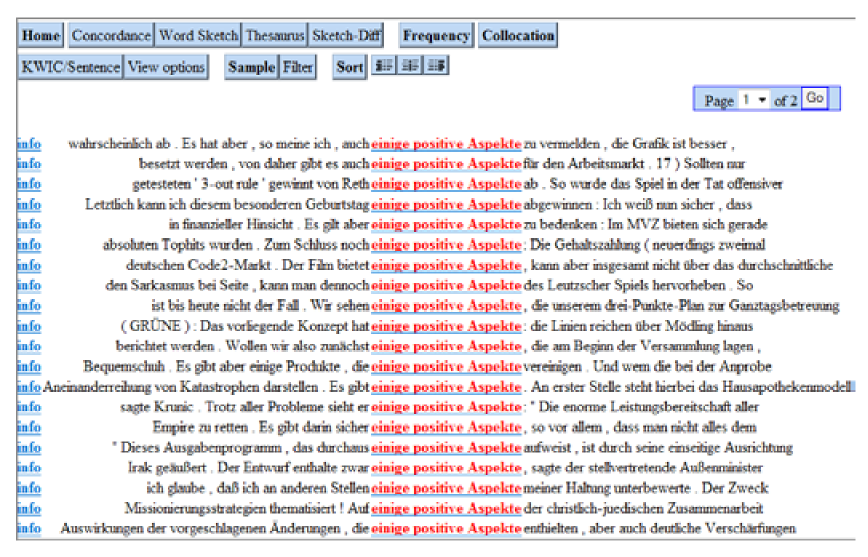



Abb. 4: dewac – Korpus Südtirol

Ist einmal die Sensibilität für diese Art von Gebrauchsbeschränkungen von Konstruktionen geweckt, werden auch andere Ausdrücke des Textes fragwürdig: *eine einzige Sprache*, *(hat) Einstellungen verbunden*, *Dank der Sprache*, *nutzen sie die Sprache* u. a. m.

Die Überprüfung dieser fragwürdigen Fälle wird von den Studierenden bereits selbstständig durchgeführt (vorteilhaft ist die Arbeit in Paaren oder in Kleingruppen). Als Ergebnis zeigt sich ein durchgehender Mangel an Idiomatizität, der in dieser Konstanz zu weiteren Fragen und Hypothesen Anlass gibt. So zur Hypothese, die mangelnde Idiomatizität stehe in kausaler Verbindung zur Isolation der Sätze als Einzelsätze ohne Textzusammenhang.

Der erfahrenen Lehrperson wird sich darüber hinaus auch die Frage nach der Entstehung der deutschen Sätze stellen, die, den Erfahrungen nach, nicht als selbstständige deutsche Sätze, sondern als Übersetzungen aus dem Italienischen entstanden sind. Die Lernenden geben auch freimütig zu, dass fast alles, was sie schreiben, aus dem Italienischen übersetzt wurde, so auch die gerade analysierte Nominalgruppe, die dem italienischen Ausdruck *alcuni aspetti positivi* entspricht.

Nachdem die Suche nach dem deutschen Ausdruck die für alle Teilnehmer unerwartete Gebrauchsbeschränkung nachgewiesen hat, wird auch der italienische Ausdruck fragwürdig. Und mit demselben Suchinstrument wird das italienische Internetkorpus aktiviert und die Suche nach der festen Wortverbindung durchgeführt, die das folgende – nicht mehr überraschende – Ergebnis zeigt:

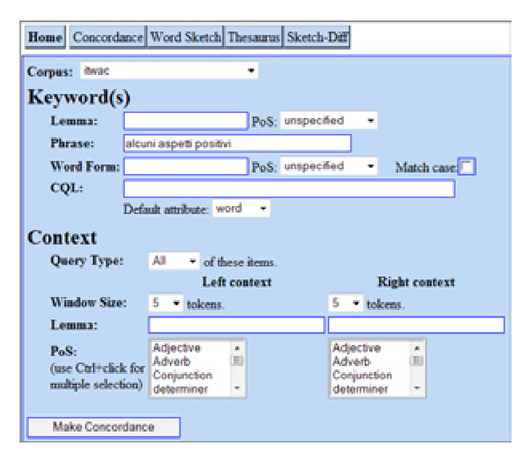



Abb. 5: itwac – Korpus Südtirol

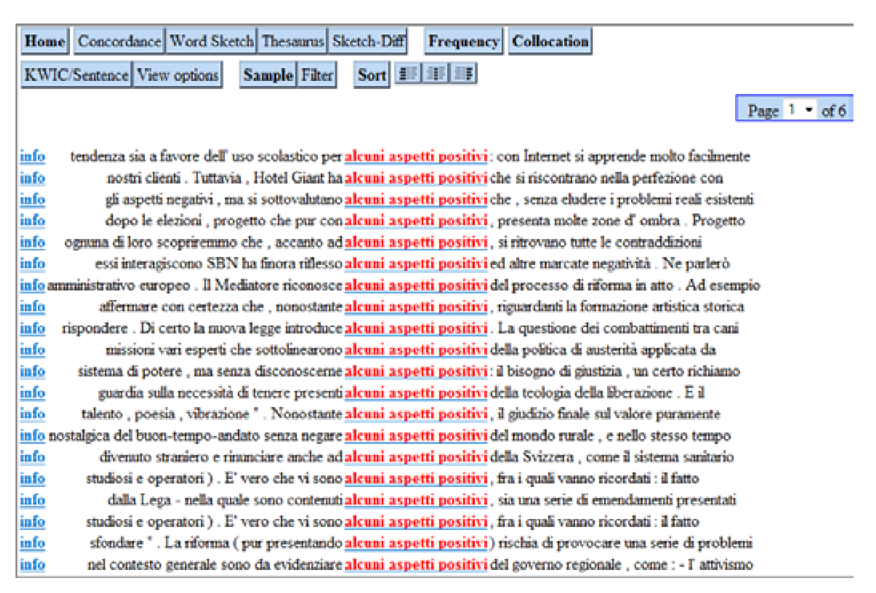



Abb. 6: itwac – Korpus Südtirol

Im Italienischen gilt also genau dieselbe Gebrauchsbeschränkung wie im Deutschen. Die Studenten haben den italienischen Ausdruck schon in der italienischen Übersetzungsvorlage *falsch* verwendet.

Wir haben ein Mosaiksteinchen für das Gesamtbild des Lernprozesses, der seinen Ausgangspunkt in italienischen Sätzen hat, die eigens dafür geschaffen werden, um als Vorlage für die Übersetzung ins Deutsche zu dienen. Die fehlende Textualität, die wir im deutschen Text diagnostiziert haben, stammt in vielen Fällen schon aus der italienischen Vorlage und entsteht durch die Isolation der italienischen Sätze als Einzelsätze, die nur als Übersetzungshilfe entstanden sind und nicht in der Absicht, einen kohärenten italienischen Text zum behandelten Thema zu verfassen.
Im konkreten Fall wurde der Grammatikfehler im ersten Satz nicht markiert. Die Korrektur des Fehlers hätte – so können wir pointiert sagen – vom Wesentlichen abgelenkt, hätte das Wesentliche noch unsichtbarer gemacht, als es ohnedies schon war.
Diese Suchstrategien könnten auch mit Studenten deutscher Muttersprache zur Stärkung der Sprachaufmerksamkeit in der Muttersprache und der Idiomatizität in der Zweitsprache und im Englischen eingesetzt werden. Auch im Englischen gelten dieselben Beschränkungen.
Viele Lernende reagieren auf Erfahrungen, wie die gerade beschriebene, sehr positiv. Sie erleben die Erkenntnis, dass die Fehlerquelle gleichsam unsichtbar in der Muttersprache steckt, als Aha-Erlebnis, das ihre Haltung zum eigenen Sprachgebrauch schlagartig ändert. Die Bereitschaft zu einer nachhaltigen Änderung im Lernverhalten entsteht. Die an mehreren Sprachen – vor allem auch an der Muttersprache – geschulte Sprachaufmerksamkeit ist ein Mittel, um auf dem neuen Weg erfolgreich weiter zu gehen.

## 2. Die Schnittstelle Input-Sprachverarbeitung

Im Abschnitt über Lernervarietäten schlägt Wolfgang Klein eine Brücke zwischen der empirischen Spracherwerbsforschung europäischer Tradition und den input-orientierten methodischen Ansätzen, die sich in den Vereinigten Staaten durchgesetzt haben. (Krashen 1988; VanPatten 2002)

Allen bisher diskutierten Theorien ist eigen, dass sie den Erwerbsprozess vorrangig unter dem Gesichtspunkt des verpassten Ziels betrachten: es gibt einen Standard, den der Lerner erreicht haben sollte - nämlich die Zielsprache -, und man bemisst den Erwerb daran, wo und warum er dieses Ziel (noch) verfehlt. Unter normativen Aspekten ist dies eine höchst verständliche Betrachtungsweise, und ein Lehrer ist gut beraten, sie einzunehmen. Sie hat auch den nicht zu unterschätzenden Vorzug der methodischen Einfachheit: man ermittelt einfach die "Fehler" des Lerners, d.h. seine Abweichungen vom Ziel, und versucht, ihr Zustandekommen zu erklären. Dies erklärt den erstaunlichen Erfolg der "Fehleranalyse" (Corder 1981) bei der Erforschung des ge-

> steuerten Zweitspracherwerbs. Es ist aber wenig wahrscheinlich, dass man auf diese Weise versteht, was im Erwerbsprozess eigentlich abläuft: dies wird davon bestimmt, wie der Lerner gewisse Fertigkeiten aus bestimmten Gründen auf einen bestimmten Input anwendet. Wenn man die immanente Gesetzlichkeit dieses Prozesses verstehen will, muss man den Lerner selbst in den Mittelpunkt rücken. (Klein 2001: 615)

Beim Lernen, und vor allem aus der Perspektive gesteuerter Lernprozesse, geht es darum, Klarheit zu gewinnen über die Fertigkeiten, die Lernende in bestimmten Momenten auf den Input anwenden, mit dem sie gerade in Kontakt gekommen sind. Der zuvor analysierte Text eines fortgeschrittenen Studierenden zeigt sehr schön, wie nicht saliente Elemente der Sprache selbst nach Jahren der Übung im Sprachgebrauch nicht korrekt gebraucht werden. So verschwindet das –n der Adjektivendung in der Nominalgruppe „Die Einführung der deutsche_ Sprache" sowohl in der gesprochenen wie auch in der geschriebenen Sprache mit großer Konstanz. Der „Fehler" ist, wie alle Lehrpersonen wissen, nur schwer korrigierbar. Die Antwort der Lernenden, „Natürlich, das N" führt nicht zu einem Kompetenzschub und zu einer nachhaltigen Befähigung, die Morphologie der Nominalgruppe aktiv anzuwenden.
Im jeweils aktualisierten Sprechakt kommen Äußerungen vor, in denen Nominalgruppen als bedeutende Elemente im Informationsfluss eine Rolle spielen. Die Nomina stellen Fixpunkte für die Intention und die kommunikative Lösung der intendierten Äußerung dar und werden mit entsprechend großer Aufmerksamkeit wahrgenommen. Der Anteil der Morphologie ist gegenüber der salienten semantischen Information immer im Hintergrund und wird immer aufs Neue „vergessen". Ein Nicht-Achten, das im Moment des Sprechens und Schreibens automatisch durch die Bevorzugung der Inhaltselemente ausgelöst wird. Auch jahrelange Vertrautheit mit der Sprache führt nicht notwendigerweise zu einer Verbesserung der Kompetenz. Die Grammatikmorpheme sind nicht-salient und bleiben bei der kognitiven Verarbeitung unberücksichtigt. Die wachsende Vertrautheit mit der Sprache, das Anwachsen des Wortschatzes gehen nicht Hand in Hand mit einer Stärkung der grammatikalischen Kompetenz der nicht salienten, im Deutschen auch durch Deakzentuierung unauffällig erscheinenden Grammatikmorpheme.
Die besondere Beachtung der Verarbeitung der sprachlichen Stimuli im Moment des Lernens ist in ein theoretisches Gerüst von Modellen kognitiver Sprachverarbeitung eingebettet. Die Forschung war in den letzten Jahrzehnten stark geprägt von einer grundlegend neuen Ausrichtung auf kognitive Problemstellungen und Zielsetzungen. Zwei große Lager haben sich gebildet, die bis heute noch kaum Tendenzen zu einem Wissensaustausch zeigen. Grundle-

gend getrennt sind die beiden Forschungsrichtungen von der Definition der kognitiven Grundlegung des Lernens von Sprachen. Auf der einen Seite die Annahme einer als konstant und unveränderlich gedachten kognitiven Grundlage in Form der Universalgrammatik, die stärker noch als der LAD diese Forschungsrichtung nachhaltig geprägt hat, und auf der anderen Seite eine Sicht der Kognition als emergente, nicht sprachspezifische Infrastruktur des Menschen, die nach allgemeinen Prinzipien der Verarbeitung funktioniert.
In beiden Positionen sind die Begriffe Sprachverarbeitung (*Processing*) zentral, sowie der strategische Ausgangspunkt beim *Input*, der die Grundlage allen impliziten und expliziten Lernens darstellt. In beiden „Schulen“ wurde, unabhängig von einander, die Erkenntnis gewonnen, dass erwachsene Lernende gegenüber dem Spracherwerb bei Kindern eine Charakteristik aufweisen, die weitreichende Folgen für den Erwerbs-, bzw. Lernprozess hat: Erwachsene erreichen die Form-Funktion-Verbindung nicht mehr spontan wie es beim frühkindlichen Spracherwerb der Fall ist.
Bill VanPatten hat, ausgehend von dieser Erkenntnis, eine erfolgreiche Forschungsrichtung im Bereich der SLA als „Processing Instruction“ begründet. Er geht über die Input-These von Krashen hinaus, da Lernende, wenn sie nicht in einer vorbereitenden Phase ihren Input für das nachfolgende Lernen optimieren, in gesteuerten Lernsituationen das Lernziel nicht erreichen, das sie im Kontakt mit verständlichem Input (auch ohne Anleitungen durch regelbasiertes Verhalten) erreichen sollten. Der Begriff des Lernens ist dabei verpönt und wird seit Krashen, nicht ohne eine gewisse ideologische Verhärtung, systematisch durch den Begriff des Spracherwerbs (acquisition) ersetzt. Lernen in Freiheit von Constraints, durch freie Verarbeitung der sprachlichen Stimuli steht in Opposition zur Vorstellung, die Krashen und VanPatten leitet, die Vorstellung eines Erwerbens der Sprache auf Grund der angeborenen kognitiven Voraussetzungen und des Setzens von Parametern bezüglich der jeweils aktuellen Einzelsprache.

Van Patten verweist darauf, dass die Lernenden dazu befähigt werden müssen, den fremdsprachlichen Text so zu be- und verarbeiten, dass sie die dem Text zugrunde liegenden Grammatikstrukturen erkennen und verstehen. Er nennt diese Art des Lehrens *processing instruction* (PI).
*Processing instruction* sieht vor, dass:

1) dem Lerner die Grammatikregel erklärt wird;
2) der Lerner auf die Schwierigkeiten aufmerksam gemacht wird, die ihm die Grammatikregel bereiten könnte, insbesondere im Falle von Unter-

schieden zwischen der Grammatikregel in der Zielsprache und der in der Muttersprache

3) der Lerner Übungen macht, bei denen er die Regel erkennen muss, um den Inhalt zu entschlüsseln.

Wie alle radikal innovativen Methoden, zeigt auch die PI eine starke Abhängigkeit von den Entstehungsbedingungen der Theorie, in diesem Fall von der Ausrichtung auf Spanisch als Fremd- oder Zweitsprache für englischsprachige Lernende. Die Position der Elemente in der linearen Abfolge der Sprachzeichen hervorzuheben ist nämlich gerade in solchen Fällen plausibel (bzw. folgenlos), wo Sprecher einer akzentzählenden Sprache mit einer stark durch die silbenzählende Struktur charakterisierten Sprache in Kontakt treten, so bei Spanisch für englischsprechende Lerner, aber nicht umgekehrt.

Bei Sprachen wie Deutsch, bei denen die lineare Struktur der Abfolge durch eine prosodisch gestaltete Informationsstruktur überlagert wird, ist die Salienz der einzelnen Satzglieder von der Akzentuierung bestimmt und nicht von der Position in der Äußerung.

Diese „Ausnahmen“ werden dem auf der Grundlage des Englischen erstellten Konzepts bei VanPatten in einer späteren Phase „hinzugefügt“. Andere Elemente der Strukturierung, die kongruent zur Ausformung von Akzentgruppen am Aufbau der Sprache beteiligt sind, bleiben hingegen unbeachtet.

So weit die allgemeine Charakterisierung dieses neuen methodischen Ansatzes, um ihm gegenüber den Ansatz des korpusgestützten Lernens abzusetzen.

Einen grundsätzlich anderen Weg geht die wissenschaftliche Schule, die Sprache als emergentes System ansieht und die ich hier an Hand der Arbeiten von Nick Ellis referieren möchte, da Ellis im Dialog zur Schule VanPattens aufgetreten ist.

Die unterschiedlichen Ansätze treffen sich in der theoretisch begründeten Grundsatzentscheidung, den Erwerbsprozess, das Erlernen der Sprache, als „ungesteuerten“ Prozess anzusehen, der nur ausgelöst und gelenkt, nicht aber wirklich methodisch „beherrscht“ werden kann. Nick Ellis, der ein umfassendes Konzept vorgelegt hat, erfasst sehr genau den kritischen Moment der Theoriebildung, die am Begriff des Input ansetzt:

> Associative learning provides the rational mechanisms for first language acquisition from input-analysis and usage, allowing just about every human being to acquire fluency in their native tongue. Yet although second language learners too are surrounded by language, not all of it ‘goes in’, and SLA is typically much less successful than L1A.
>
> This is Corder’s distinction between input, the available target language, and intake, that subset of input that actually gets in and which the learner utilizes in some way (Corder 1967). Does this mean

> that SLA cannot be understood according to the general principles of associative learning that underpin other aspects of human cognition? If L1A is rational, is L2A fundamentally irrational?
> Children learning their native language only acquire the meanings of temporal adverbs quite late in development. But second language learners already know about adverbs from their first language experience, and adverbs are both salient and reliable in their communicative functions, while tense markers are neither. Thus, the second language expression of temporal reference begins with a phase where reference is established by adverbials alone (Bardovi-Harlig 1992; Meisel 1987), and the grammatical expression of tense and aspect thereafter emerges only slowly, if at all. (Bardovi-Harlig, 2000)

Ungesteuertes Lernen bedarf der Aufmerksamkeit als allgemeiner Haltung gegenüber den im Kontakt mit der neuen Sprache empfangenen Vorlagen und hat im Falle des Zweitspracherwerbs mit echten Lernbehinderungen zu rechnen, die darin bestehen, dass die in der Erstsprache erworbenen metasprachlichen Kenntnisse gerade grammatikalisch saliente Elemente der neuen Sprache als redundante Angaben ausweisen, die der Aufmerksamkeit der Lernenden entzogen werden.

Das kommunikativ und inhaltsbezogene Lernen – auch in Fach-Sprach-Unterrichtsmodellen – muss in dieser Perspektive durch Übungen ergänzt werden, die gerade die nicht salienten Merkmale der Sprache unauffällig aber effizient in den Vordergrund rücken. Das Hauptaugenmerk liegt nicht so sehr auf der Form der Darstellung, also der Qualität, als vielmehr auf der Quantität der dargestellten Formen.

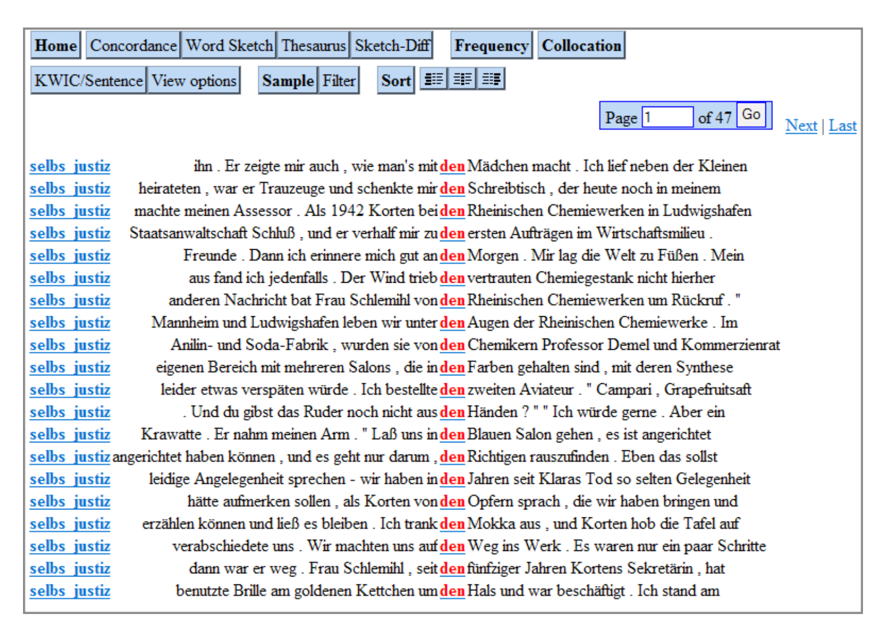
Das Erlernen einer neuen Sprache in der Schule braucht nicht so sehr neue methodische Inspiration als vielmehr einen quantitativen Schub, der von den punktuellen Momenten des Sprachkontakts zu einer Begegnung mit der neuen Sprache führt, die dem sogenannten „Sprachbad“ der natürlichen Spracherwerbssituation vergleichbar ist. Ein Mehr an verständlichem Input, das zu erreichen allerdings jenseits der Unterrichtsstunden liegt (Drumbl 2008a). Der Unterricht hat sich diesen quantitativen Anforderungen zu stellen, und sich ihnen gegebenenfalls auch unterzuordnen. Genau an dieser Stelle, wo Anforderungen der Qualität und Quantität zusammentreffen, haben gesteuerte Übungen auf der Grundlage von Korpusrecherchen ihren Platz.

Im Sprachsyllabus für Anfänger ist in unseren Kursen die Lektüre des Romans „Selbs Justiz“ von Bernhard Schlink und Harald Popp vorgeschrieben (in Ausnahmefällen begnügen wir uns mit dem ersten Teil, mit seiner abgeschlossenen Handlung). Der Roman ist als Korpus aufbereitet und kann mit der auch bei den anderen Quellen benutzten Suchsoftware durchsucht wer-

den. Die Beschränkung auf einen kurzen Text lässt folgende untypische Charakteristiken der Korpusarbeit hervortreten.

1. Alle Belege sind brauchbar.
2. Alle Belege werden vom Lernenden beim Lesen des Buches wiedergefunden.
3. Alle Belege gehören zur Basisgrammatik, die die Studierenden auf jeden Fall lernen „wollen".

Sind diese Voraussetzungen gegeben, kann schon die als *Wordform* eingegebene Suche nach dem Artikel „den" von großem didaktischen Nutzen sein. Die Belege betreffen sowohl den Akkusativ Singular als auch den Dativ Plural. Jeder Beleg weist redundante Merkmale auf, um den Singular vom Plural zu unterscheiden, und die Anwesenheit oder Abwesenheit von Präpositionen gibt einen klaren Hinweis auf den Kasus.




| | | | |
|---|---|---|---|
| selbs_justiz | ihn . Er zeigte mir auch , wie man's mit | den | Mädchen macht . Ich lief neben der Kleinen |
| selbs_justiz | heirateten , war er Trauzeuge und schenkte mir | den | Schreibtisch , der heute noch in meinem |
| selbs_justiz | machte meinen Assessor . Als 1942 Korten bei | den | Rheinischen Chemiewerken in Ludwigshafen |
| selbs_justiz | Staatsanwaltschaft Schluß , und er verhalf mir zu | den | ersten Aufträgen im Wirtschaftsmilieu . |
| selbs_justiz | Freunde . Dann ich erinnere mich gut an | den | Morgen . Mir lag die Welt zu Füßen . Mein |
| selbs_justiz | aus fand ich jedenfalls . Der Wind trieb | den | vertrauten Chemiegestank nicht hierher |
| selbs_justiz | anderen Nachricht bat Frau Schlemihl von | den | Rheinischen Chemiewerken um Rückruf . " |
| selbs_justiz | Mannheim und Ludwigshafen leben wir unter | den | Augen der Rheinischen Chemiewerke . Im |
| selbs_justiz | Anilin- und Soda-Fabrik , wurden sie von | den | Chemikern Professor Demel und Kommerzienrat |
| selbs_justiz | eigenen Bereich mit mehreren Salons , die in | den | Farben gehalten sind , mit deren Synthese |
| selbs_justiz | leider etwas verspäten würde . Ich bestellte | den | zweiten Aviateur . " Campari , Grapefruitsaft |
| selbs_justiz | . Und du gibst das Ruder noch nicht aus | den | Händen ? " " Ich würde gerne . Aber ein |
| selbs_justiz | Krawatte . Er nahm meinen Arm . " Laß uns in | den | Blauen Salon gehen , es ist angerichtet |
| selbs_justiz | angerichtet haben können , und es geht nur darum , | den | Richtigen rauszufinden . Eben das sollst |
| selbs_justiz | leidige Angelegenheit sprechen - wir haben in | den | Jahren seit Klaras Tod so selten Gelegenheit |
| selbs_justiz | hätte aufmerken sollen , als Korten von | den | Opfern sprach , die wir haben bringen und |
| selbs_justiz | erzählen können und ließ es bleiben . Ich trank | den | Mokka aus , und Korten hob die Tafel auf |
| selbs_justiz | verabschiedete uns . Wir machten uns auf | den | Weg ins Werk . Es waren nur ein paar Schritte |
| selbs_justiz | dann war er weg . Frau Schlemihl , seit | den | fünfziger Jahren Kortens Sekretärin , hat |
| selbs_justiz | benutzte Brille am goldenen Kettchen um | den | Hals und war beschäftigt . Ich stand am |

Abb. 7: http://search.korpus-suedtirol.it:8091/

Die Aufgabe lautet einfach, die Belege nach Mustern zu sortieren. Das geschieht im gesteuerten Unterricht in Zweiergruppen. Je zwei Lernende arbeiten an einem PC und ordnen die Belege in einem gemeinsamen Sprechakt durch halblautes Vorlesen der KWIC-Stellen und Kommentieren („Typus eins").

Die Belege werden in der Reihenfolge der ersten Seite gegeben:

| | | | |
|---|---|---|---|
| 1 | Er zeigte mir auch, wie man's mit | den | Mädchen macht. Ich lief neben der Kleinen |
| 2 | heirateten, war er Trauzeuge und schenkte mir | den | Schreibtisch, der heute noch in meinem |

Die unvorhersehbaren Bruchteile von Sätzen in den einzelnen Zeilen provozieren gleichsam inferenziell forschendes Weiterdenken und „Weiterlesen" über die Bruchstellen hinweg. Da der erweiterte Kontext durch einen Klick ohne Verzögerung aufgerufen werden kann, nehmen Studierende diese Option in ausgewählten Fällen spontan wahr. Die unvollständigen Sätze mit ihren salienten Mustern können daher schon bei mäßig fortgeschrittenen Lernenden zur Schulung der Inferenz genutzt werden, mit selbstgesteuerter Überprüfung der Ergebnisse des inferentiellen Erprobens. Aber auch ohne Verständnis des Kontextes wird bei diesem Beleg die Grammatik der Erstfeld-Besetzung transparent.

| | | | |
|---|---|---|---|
| 3 | machte meinen Assessor. Als 1942 Korten bei | den | Rheinischen Chemiewerken anfing |
| 4 | und er verhalf mir zu | den | ersten Aufträgen im Wirtschafts-milieu |
| 5 | Freunde. Dann ich erinnere mich gut an | den | Morgen. Mir lag die Welt zu Füßen. Mein |

Beleg 5 mit einem OCR-Fehler, „Dann" an Stelle von „Denn", der vom Korrekturprogramm nicht erfasst wurde.

| | | | |
|---|---|---|---|
| 6 | aus fand ich jedenfalls. Der Wind trieb | den | vertrauten Chemiegestank nicht hierher |
| 7 | anderen Nachricht bat Frau Schlemihl von | den | Rheinischen Chemiewerken um Rückruf |
| 8 | Mannheim und Ludwigshafen leben wir unter | den | Augen der Rheinischen Chemie-werke. Im |
| 9 | Anilin- und Soda-Fabrik , wurden sie von | den | Chemikern Professor Demel und Kommerzienrat |
| 10 | eigenen Bereich mit mehreren Salons, die in | den | Farben gehalten sind, mit deren Synthese |

| 11 | leider etwas verspäten würde . Ich bestellte | den | zweiten Aviateur. „Campari, Grapefruitsaft |
|---|---|---|---|
| 12 | . Und du gibst das Ruder noch nicht aus | den | Händen?““Ich würde gerne. Aber ein |
| 13 | Krawatte. Er nahm meinen Arm . „ Laß uns in | den | Blauen Salon gehen, es ist ange-richtet |
| 14 | angerichtet haben können, und es geht nur darum, | den | Richtigen rauszufinden. Eben das sollst |
| 15 | leidige Angelegenheit sprechen – wir haben in | den | Jahren seit Klaras Tod so selten Gelegenheit |
| 16 | Hätte aufmerken sollen, als Korten von | den | Opfern sprach, die wir haben bringen und |
| 17 | Erzählen können und ließ es bleiben. Ich trank | den | Mokka aus, und Korten hob die Tafel auf |
| 18 | verabschiedete uns . Wir machten uns auf | den | Weg ins Werk . Es waren nur ein paar Schritte |
| 19 | dann war er weg. Frau Schlemihl , seit | den | fünfziger Jahren Kortens |
| 20 | benutzte Brille am goldenen Kettchen um | den | Hals und war beschäftigt. Ich stand am |

Die grammatikalischen Lernobjekte sind nach Mustern sortiert, sie erscheinen aber in aleatorischer Abfolge in unterschiedlichen Kontexten: die Grammatikmorpheme werden daher in einer natürlichen sprachlichen Umgebung wahrgenommen, nämlich als nicht saliente Elemente, sie treten aber gehäuft auf und werden unbewusst mitverarbeitet.

Die dabei aktivierten kognitiven Fähigkeiten werden bei der Arbeit mit den Ergebnissen der Korpussuche im Detail erkennbar. Das rasche Aussortieren von nicht pertinenten Belegen im Hinblick auf ein Muster, das der Lernende selbst als Ausgangspunkt seiner Suche gewählt hat. Die dafür notwendige Aufmerksamkeit, die auf formale Aspekte der Sprache in Mustern und Gruppen, auf die Gliederung nach Mehr-Wort-Konstruktionen ausgerichtet ist, kann den Blick schärfen für weitere rekurrente Strukturen, die in den Belegen zu erkennen sind. Der nach formalen und zugleich funktionalen Kriterien unterscheidende Blick wird durch Üben daran gewöhnt, immer umfassendere Elemente des Kontextes in die Analyse einzubeziehen und gewinnt Sicherheit im Suchen, im Erkennen von Regelmäßigkeiten, von Mustern und von Differenzen.

Diese kognitive Aktivität sind keine Leistungen, die nur von fortgeschrittenen Lernenden abverlangt werden können. In dieser Perspektive gewinnen sogar anscheinend triviale Recherchen wie die nach einzelnen Präpositionen Kontur. Die Suche nach dem Lemma *auf* bietet schon auf der ersten Seite eine Fülle von Belegen, die zur Musterdifferenzierung einladen. Auch Anfänger erkennen an rekurrenten Morphemen wie *der dem* grammatikalische Formen und Formeln.


Home | Concordance | Word Sketch | Thesaurus | Sketch-Diff | Frequency | Collocation
KWIC/Sentence | View options | Sample | Filter | Sort
Page 1 of 31 Go   Next | Last

| Source | Left context | KWIC | Right context |
|---|---|---|---|
| selbs_justiz | Am Anfang habe ich ihn beneidet . Das war | auf | der Schule , auf dem Friedrich-Wilhelm-Gymnasium |
| selbs_justiz | ihn beneidet . Das war auf der Schule , | auf | dem Friedrich-Wilhelm-Gymnasium in Berlin |
| selbs_justiz | Friedrich-Wilhelm-Gymnasium in Berlin . Ich trug Vaters Anzüge | auf | , hatte keine Freunde und kam nicht am |
| selbs_justiz | heute selten dran . Mein Beruf hält mich | auf | Trab , und wenn ich abends noch kurz ins |
| selbs_justiz | noch kurz ins Büro schaue , türmen sich | auf | dem Schreibtisch keine Akten . Nur der |
| selbs_justiz | Besuche zu Weihnachten und zum Geburtstag | auf | . Wir verkehren in verschiedenen Kreisen |
| selbs_justiz | Schokoladenhörnchen , und ich frühstückte draußen | auf | dem Gehsteig in der Sonne . Eine junge |
| selbs_justiz | hübscher , und ich stellte meine Einwegtasse | auf | das Schaufenstersims und ging hinter ihr |
| selbs_justiz | Büro in der Augusta-Anlage . Ich bin stolz | auf | mein Büro . In Tür und Schaufenster des |
| selbs_justiz | Lettern : Gerhard Selb Private Ermittlungen | Auf | dem Anrufbeantworter waren zwei Anrufe |
| selbs_justiz | Sie sah mich mitfühlend an . " Sie müssen | auf | Herrn Generaldirektor warten ? " Ich hatte |
| selbs_justiz | hängen würde . " Mein lieber Selb " , kam er | auf | mich zu . Klein und drahtig , mit wachsamen |
| selbs_justiz | Trinken ? " Er warf einen zweifelnden Blick | auf | mein Päckchen Sweet Afton . " Bringen Sie |
| selbs_justiz | Rohkostsalat essen , welcher Mitarbeiter gerade | auf | unserem Tennisplatz trainiert , die intakten |
| selbs_justiz | ihn zu verachten , um so stolzer war er | auf | seine Enkelkinder . Marion war in die Studienstiftung |
| selbs_justiz | den Mokka aus , und Korten hob die Tafel | auf | . Der Chef des Kasinos verabschiedete uns |
| selbs_justiz | Kasinos verabschiedete uns . Wir machten uns | auf | den Weg ins Werk . Es waren nur ein paar |
| selbs_justiz | ein Gewirr von Türmen , Hallen und Rohren | auf | den Handelshafen und das dunstblasse Mannheim |
| selbs_justiz | zurückfordert . " In ihrem Vorzimmer drückte sie | auf | den Knopf der Sprechanlage . " Herr Direktor |
| selbs_justiz | flink und der Händedruck nicht fester . | Auf | seinem Schreibtisch lag ein dicker Ordner |

Fazit: Der eigentliche Gewinn der systematischen Korpusarbeit im Sprachunterricht liegt also vielleicht gerade darin, dass ganz klar formulierte Ziele vorgegeben werden, die zu klar bestimmten Übungsanordnungen führen. Korpusrecherchen sind Übungen, die relevante Ziele verfolgen. Die Arbeit mit Korpora ist zugleich aufmerksames Beobachten des Sprachgebrauchs, Lernen und Üben. Der Akzent liegt auf dem Üben, der unbekannten Größe allen Lernens.

## Literaturverzeichnis


Abel, Andrea / Anstein, Stefanie / Petrakis, Stefanos (2009): *Die Initiative Korpus Südtirol*. In: Linguistik online 38, 2/2009.

Adolphs, Svenja / Durow, Valerie (2004): *Social-cultural integration and the development of formulaic sequences.* In: Schmitt, Norbert (Hg.): Formulaic Sequencies. Acquisition, processing and use. Amsterdam-Philadelphia: John Benjamins, 107-126.

Ágel, Vilmos (2004): *Phraseologismen als (valenz)syntaktischer Normalfall*. In: Steyer, K. (Hg.): Wortverbindungen - mehr oder weniger fest. IDS-Jahrbuch 2003. Berlin, New York, 65-86

Bardovi-Harlig, Kathleeen (1992): *The use of adverbials and natural order in the development of temporal expression*. IRAL XXX: 299–320.

Bardovi-Harlig, Kathleen (2000): *Tense and Aspect in Second Language Acquisition: Form, meaning, and use*. Oxford: Blackwell.

Baroni, Marco / Kilgariff, Adam (2006): *Large linguistically-processed Web corpora for multiple languages*. http://clic.cimec.unitn.it/marco/publications/eacl2006/dewac_eacl.pdf

Bates, Elizabeth / Goodman, Judith C. (1997): *On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-Time Processing*. In: Language and Cognitive Processes, 12 (5/6), 507-584.

Bates, Elizabeth / Elman, Jeffrey (1996): *Learning Rediscovered*. In: Science, vol. 274, 13. Dezember 1996.

Belica, Cyril (2001): *Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. http://corpora.ids-mannheim.de/ccdb/.

Bolla, Elisabetta / Drumbl, Hans (2009): *Sprachen, Kulturen, Grenzen. Teilhabe an der Sprachgemeinschaft: eine Schlüssel zur Zweitsprache*. Meran: Alpha&Beta.

Chomsky, Noam/Lasnik, Howard (1993): *The Theory of Principles and Parameters*. In: Joachim Jacobs et al. (Hg.), Syntax, HSK 9.1., Berlin, 506-569.

Corder, S. Pit (1981): *Error analysis and interlanguage.* Oxford: Oxford University Press.

Decoo, Wilfried (2000): *On the mortality of language learning methods*, given as the James L. Barker lecture on November 8th 2001 at Brigham Young University.

Dörnyei, Zoltán/Durow, Valerie/Zahran, Khawla (2004): *Individul differences and their effects on formulaic sequence acquisition.* In: Schmitt, Norbert (Hg.): Formulaic Sequencies. Acquisition, processing and use. Amsterdam-Philadelphia: John Benjamins, 87-106.

Drumbl, Johann (2002): *Aber das Eigene muss so gut gelernt seyn wie das Fremde*. In: Portmann-Tselikas, Paul R. / Schmölzer-Eibinger, Sabine (Hgg.): Theorie und Praxis. Österreichische Beiträge zu Deutsch als Fremdsprache. Textkompetenz. Neue Perspektiven für das Lernen und Lehren. Innsbruck: StudienVerlag, 247-260.

Drumbl, Johann: (2003): *Das Sprachen-Portal. Inferenz und Spracherwerb in mehrsprachiger Lernumgebung*. Meran: AlphaBeta.

Drumbl, Johann (2007): *Sprachunterricht und Ausbildung zur Mehrsprachigkeit. Bilder, Gedanken, Erinnerungen*. In: Schmölzer-Eibinger, Sabine/Weidacher, Georg (Hgg.): Textkompetenz. Eine Schlüsselkompetenz und ihre Vermittlung. Tübingen: Narr, 299-310.

Drumbl, Hans (2008a): *Leben mit Sprachen*. In: ff Südtiroler Wochenmagazin Nr. 04, 38-39.

Drumbl Hans (2008b): *Auf dem Weg zur Sprachaufmerksamkeit*. Academia Nr. 46, http://www.eurac.edu/academia/44/academia44.pdf, Wissenschaftsmagazin der Europäischen Akademie Bozen (Interview).

Drumbl, Hans: *Kognition und Lernen: neue Erkenntnisse zur Wortschatzarbeit*, Studienverlag, Innsbruck, Wien (im Druck).

Drumbl, Hans: *Lingue da vivere.* In: Prader, Luis Thomas / Dal Negro, Silvia / Civegna, Klaus (Hgg.): Deutsch als Zweit- bzw. Ortssprache an den Schulen der deutschsprachigen Minderheiten in Italien bzw. an den italienischen Schulen Südtirols. Tagung, Brixen, 7./8. März 2008 (im Druck).

Duffner, Rudolf/Näf, Anton (2006): *Digitale Textdatenbanken im Vergleich*. http://www.linguistik-online.de/28_06/duffnerNaef.html.

Ellis, Nick C. (2006*): Cognitive Perspectives on SLA: The Associativ-Cognitive CREED.* In: Bardovi-Harlig, Kathleen / Dörnyei, Zoltà (Hgg.): Themes in SLA Research. AILA Review, Volume 19, Amsterdam-Philadelphia.

Evert, Stefan (2005):*The Statistics of Word Cooccurrences: Word Pairs and Collocations.* http://elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf (Dissertation an der Universität Stuttgart).

Feilke, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt a. M.: Suhrkamp.

Feilke, Helmuth (1998): *Idiomatische Prägung*. In: Barz, Irmhild/Öhlschläger, Günther (Hg.): Zwischen Grammatik und Lexikon. Tübingen: 69–80. (= Linguistische Arbeiten 390).

Fraas, Claudia (2001), *Usuelle Wortverbindungen als sprachliche Manifestation von Bedeutungswissen. Theoretische Begründung, methodischer Ansatz und empirische Befunde*. In: Nikula, Henrik / Drescher, Robert (Hgg.): Lexikon und Text, Vaasa, 41-66.

Gasser, Michael (1994): *Modularity in a connectionist Model of Morphology Acquisition*, http://www.aclweb.org/anthology/C/C94/C94-1034.pdf (25.01.2008).

Gasser, Michael/Smith, Linda B. (1996): *Comparison, Categorization, and Perceptual Dimensions: A Connectionist Model of the Development of the Notion of Sameness*, http://www.cs.indiana.edu/pub/gasser/cogsci91.pdf (20.01.2008).

Geyken, Alexander (2004): *Korpora als Korrektiv für einsprachige Wörterbücher*. In: Lili. Zeitschrift für Literaturwissenschaft und Linguistik, 136, 72-100.

Gries, Stefan Th.: *Useful statistics for corpus linguistics*, http://www.linguistics.ucsb.edu/faculty/stgries/research/UsefulStatsForCorpLing.pdf

Goldberg, Adele (2006): *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Hauser, Marc D. / Chomsky, Noam / Fitch, W.Tecumseh (2002): *The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?* http://www.sciencemag.org/cgi/reprint/298/5598/1569.pdf

Klein, Wolfgang (2001): *Typen und Konzepte des Spracherwerbs.* In: Götze, Lutz /Helbig, Gerhard / Krumm, Hans-Jürgen (Hgg.): Deutsch als Fremdsprache, 1. Berlin-New York: de Gruyter, 604-616.

Kilgariff, Adam/Grefenstette Gregory (2003): *Web as Corpus: Introduction to the Special Issue*. In: http://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf.

Kilgarriff, Adam/Rychly, Pavel/Smrz Pavel/Tugwell David (2004): *The Sketch Engine.* In: *http://www.kilgarriff.co.uk/Publications/2004-KilgRychlySmrzTugwell-SkEEuralex.rtf*

Krashen, Stephen D. (1988): *Second Language Acquisition and Second Language Learning.* Prentice-Hall International.

Liebermann, Philipp (2000): *Human Language and Our Reptilian Brain. The Subcortical Bases of Speech, Syntax and Thought*. Cambridge MA: Harvard University Press.

Ludewig, Petra (2005): *Korpusbasiertes Kollokationslernen. Computer-Assisted Language Learning als prototypisches Anwendungsszenario der Computerlinguistik* (= Sprache, Sprechen und Computer 9), Frankfurt am Main: Lang.

McCarthy, Michael (2006): *Exploration in Corpus Linguistics*. Cambridge: University Press.

Meisel, Jürgen (1987): *Reference to past events and actions in the development of natural second language acquisition*. In: C. Pfaff (Hg.): First and Second Language Acquisition, New York NY: Newbury House.

Müller, Stefan (2006): *Qualitative Korpusanalyse für die Grammatiktheorie: Introspektion vs. Korpora.* In: Zifonun, Gisela / Kallmeyer, Werner (Hgg.): IDS-Jahrbuch 2006.

Perkuhn, Rainer / Belica, Cyril / Al-Wadi, Doris / Lauer, Meike / Steyer, Kathrin / Weiss, Christian (2005): *Korpustechnologie am Institut für Deutsche Sprache.* In: Schwitalla, Johannes / Wegstein, Werner (Hgg.): Korpuslinguistik deutsch: synchron - diachron - kontrastiv. Würzburger Kolloquium 20. - 23.3. 2003. Tübingen: Niemeyer, 57-70.

Pinker, Steven / Jackendoff, Ray (2005): *What's special about the human language faculty?* In: Cognition, 95(2), 201-236.

Pinker, Steven / Jackendoff, Ray (2005): *The nature of the language faculty and its implications for evolution of language* (Reply to Fitch, Hauser & Chomsky). In: Cognition, 97(2), 211-225.

Richardson, Jack C. (2008): *Moving beyond the Plateau. From Intermediate to Advanced Levels in Language Learning*. Cambridge: University Press.

Schmitt, Norbert / Dörnyei, Zoltán / Adolphs, Svenja / Durow, Valerie (2004): In: Schmitt, Norbert: Formulaic Sequencies. Acquisition, processing and use. Amsterdam-Philadelphia: John Benjamins, 55-86.

Sharoff, Serge (2005): *Creating general-purpose corpora using automated search engine queries*. http://wackybook.sslmit.unibo.it/pdfs/sharoff.pdf

Sheen, Ron (2007): *Processing Instruction.* http://eltj.oxfordjournals.org/content/61/2/161.full

Sinclair, John (1997): *Corpus, concordance, collocation* (= Describing English Language XVIII), Oxford: Oxford University Press.

Steyer, Kathrin (2004): *Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven*. In: Steyer, Kathrin (Hg.): Wortverbindungen - mehr oder weniger fest (= Jahrbuch des Instituts für Deutsche Sprache 2003). Berlin-New York: de Gruyter, 87-116.

Tomasello, Michael (2003): *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard University Press.

VanPatten, Bill (2002): *Processing Instruction: An Update*. In: Language Learning, 52, 755-803.

Wray, Alison (2000): *Formulaic Sequences in Second Language Teaching: Principle and Practice,* Applied Linguistics 21/4, 463-489.

Wray, Alison/ Perkins, Michael R. (2000): *The Functions of Formulaic Language: An Integrated Model*. In: Language and Communication 20,1-28.

Wray, Alison (2002): *Formulaic language and the lexicon,* Cambridge, UK: Cambridge University Press.

Zanin, Renata (2007): *Korpusinstrumente für Deutsch als Zweitsprache.* In: Portmann-Tselikas, Paul R./ Schmölzer-Eibinger, Sabine (Hgg.): Aufgaben. Wien, Innsbruck: Studien-Verlag.

Zanin, Renata (2008): *Parlare come gli altri. L'apprendimento del tedesco seconda lingua nella Provincia Autonoma di Bolzano*. In Pegadogia più Didattica, Baldacci, Dozza, Frabboni, Pinto Minerva. Gardolo: Erickson.

## Internetadressen

Korpus Südtirol: http://www.korpus-suedtirol.it/suche_in_den_korpora_de.htm
Digitales Wörterbuch der DeutschenSprache: http://www.dwds.de/cgi-bin/rest/login-start

# Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora

**Alexander Geyken – Berlin-Brandenburgische Akademie der Wissenschaften**

## Abstract
Das DWDS-Wortprofil ist das Ergebnis einer automatischen syntaktischen und statistischen Analyse sehr großer Korpora. Es liefert einen kompakten Überblick über die statistisch signifikanten syntagmatischen Beziehungen eines Wortes. Beispiele dieser sogenannten syntaktischen Relationen sind Attribut-Nomen Verbindungen wie *schöne Bescherung* oder Verb-Objekt Beziehungen wie *Flasche entkorken*. Die Darstellung der Relationen erfolgt in Form einer Schlagwortwolke oder in Tabellenform. Die Berechnung des DWDS-Wortprofils erfolgt in drei Etappen: Festlegung der zu extrahierenden syntaktischen Relationstypen, Extraktion der Relationen mittels einer automatischen syntaktischen Analyse und Bewertung der statistischen Signifikanz der extrahierten Relationen. Der derzeitige Prototyp des DWDS-Wortprofils beruht auf einer Mischung eines Referenz- und eines Zeitungskorpus, dem DWDS-Kernkorpus und dem ZEIT-Archiv (1946–2009), und hat eine Gesamtgröße von 500 Millionen laufenden Textwörtern. Aus dem Korpus wurden etwa 90.000 Lemmata mit 2.000.000 Relationen extrahiert. Erste Auswertungen zeigen, dass bereits der gegenwärtige Prototyp eine große Reichhaltigkeit von Wortverbindungen enthält, die ihn auch im Vergleich mit großen einsprachigen Wörterbüchern interessant machen. Für den an der Textproduktion interessierten Nutzer bietet das Wortprofil aber einen weit über das gedruckte Wörterbuch hinausgehenden Mehrwert: Jede Relation des Wortprofils ist direkt mit den dazugehörigen Satzkontexten im Korpus verknüpft. Der Nutzer erhält somit unmittelbar einen Überblick über die weiteren semantischen und pragmatischen Kontexte in authentischen Texten. Dies kann gerade bei sprachlichen Unsicherheiten von erheblicher Hilfe sein. Das DWDS-Wortprofil ist über die Web-Plattform www.dwds.de abfragbar.

## 1. Einleitung
Elektronische Textkorpora werden seit über 20 Jahren in der Wortschatzforschung und für das Sprachlernen eingesetzt (z.B. Sinclair 1991, Braun et al. 2006). Beispiele aus Korpora haben dabei im Unterschied zu Kompetenzbeispielen den Vorzug, dass sie die verschiedenen Verwendungen des Wortes auf

der Grundlage realer Textbeispiele widerspiegeln. In der Regel basieren elektronische Plattformen, die diese Korpora analysieren, bei der Anzeige der Suchergebnisse auf Konkordanzen (s. Abb. 1). Damit können die Gebrauchskontexte eines Wortes auf übersichtliche Weise dargestellt werden. Mit wachsender Größe der Korpora - gängig sind mittlerweile Korpora einer Größe von mehreren hundert Millionen bis hin zu einigen Milliarden Textwörtern - steigt nicht nur der Variantenreichtum der Gebrauchskontexte, sondern auch die Anzahl der Belege für ein Suchwort bzw. einen Suchausdruck. In der Regel sind daher die Konkordanzanzahlen für eine Wortsuche zu groß für eine manuelle Auswertung. Beispielsweise haben die Wörter *Feindbild* oder *grau* im Archiv der ZEIT (1946–2009) Frequenzzahlen von 2364 bzw. 21656 Treffern (s. www.dwds.de).



Die ZEIT & ZEIT Online

Treffer: 2364 | 1 2 3 4 >>|

| | | | | |
|---|---|---|---|---|
| 1 | 2009 | ... dass Ihnen nach dem Abgang von George W. Bush das | Feindbild | fehlt? Goebel: Nein. |
| 2 | 2009 | ...nso wenig überraschen, schließlich sind Feinde auf | Feindbilder | angewiesen, und umgekehrt fühlte sich auch Ahmadin... |
| 3 | 2009 | ...; heute dient sie dem Anti-Drogen-Krieg, dem neuen | Feindbild | mit dem das Pentagon seine Militärpräsenz in Latei... |
| 4 | 2009 | ...ner Zeit, in der die » reiche deutsche Witwe « als | Feindbild | fungiert. Das Gespräch vor dem Werk kommt immer... |
| 5 | 2009 | ...Spiegel. Die Scharfmacher in Teheran verlieren ein | Feindbild | Man hat Barack Obama einen Präsidenten der soft po... |
| 6 | 2009 | ...rlich den Scharfmachern in Teheran, die gerade ein | Feindbild | verlieren. www.zeit.de/audio Diesen Artikel fin... |
| 7 | 2009 | ... Natürlich nutzt die Regierung den Westen auch als | Feindbild | , um sich intern zu stärken. Aber in den vergan... |
| 8 | 2009 | ...ttlerweile nicht mehr von Verschwörungsformeln und | Feindbildern | , die ebenso gut von rechtsaußen kommen könnten. ... |
| 9 | 2009 | ...en. Auch das ist neu in Lateinamerika: Das alte | Feindbild | USA dient unter Barack Obama nicht mehr zum Sünden... |
| 10 | 2009 | ... und blüht der Mythos vom freien Internet, und die | Feindbilder | sind klar. Auf der einen Seite stehen die » kon... |
| 11 | 2009 | ...öht, weil das Regime von dem Fortfall des bequemen | Feindbildes | vom Großen Satan Amerika eine Schwächung des Rückh... |
| 12 | 2009 | ...itik Von Michael Thumann Konservative beharren auf | Feindbild | Ahmadineschad Die Verteuflung Irans ist vorbei, Ba... |
| 13 | 2009 | ...Israel - konservative Konservative beharren auf | Feindbild | Ahmadineschad Iran-Politik Von Michael Thumann ... |
| 14 | 2009 | ...ud Ahmadineschad, der so bereitwillig das perfekte | Feindbild | abgibt mit seiner schlechten Politik, seinem schle... |
| 15 | 2009 | ... war vor allem das weithin geteilte dämonisierende | Feindbild | . Hetzer «, » Unterdrücker «, » Henker «: In de... |
| 16 | 2009 | ...der gelang es, nach diesem Kulturschock Europa als | Feindbild | zu porträtieren. ÖVP und SPÖ sind mitverantwort... |
| 17 | 2009 | ...ezeichnet. Politiker, die sich Manager als neue | Feindbilder | aussuchen, vergessen, dass Eliten in anständiger W... |
| 18 | 2009 | ...ise heute), das Angstschüren vor Schwarz-Gelb, das | Feindbild | (Paul Kirchhof/Karl-Theodor zu Guttenberg), die Hä... |
| 19 | 2009 | ...in seinem Leben verdienen kann, liefern ein klares | Feindbild | . Zumal sie sich in Netzwerken gegenseitig stüt... |
| 20 | 2009 | ...ans: Es sind unsere. Nur die Konstruktionen von | Feindbildern | , ob sie sich nun gegen Juden, Muslime oder Amerik... |

Treffer pro Seite: 10 20 50 100 500

Abb. 1: Konkordanzzeilen für das Substantiv Feindbild

In Abb. 1 sind die ersten 20 der 2364 Konkordanzzeilen für das Substantiv *Feindbild* dargestellt. Das Wort *Feindbild* kommt hier in sehr unterschiedlichen syntagmatischen Kontexten vor. Bei der Frage jedoch, ob und welche dieser Kontexte typisch sind, geben Konkordanzzeilen keine unmittelbare Hilfe. Wie kann man beispielsweise durch das Lesen von Konkordanzzeilen herausfinden, mit welchen Adjektiven das Substantiv *Feindbild* typischerweise verwendet wird oder umgekehrt, ob bzw. wie gebräuchlich in Abbildung 1 die Adjektive *perfekt* (Treffer 14), *dämonisierend* (15), *neu* (17) oder *klar* (19) für das Substantiv *Feindbild* sind? Ähnliche Fragen stellen sich für verbale Konstruktionen wie *auf einem Feindbild beharren* (12,13), *auf Feindbilder angewiesen*

*sein* (2), oder *ein Feindbild verlieren* (5) bzw. *ein Feindbild abgeben* (14). Um diese Fragen „per Kopf" zu beantworten, müsste man die über 2300 Konkordanzzeilen durchlesen und Strichlisten über die entsprechenden Konstruktionen führen. Es ist unmittelbar einleuchtend, dass die Beantwortung dieser Fragen bei noch häufigeren Wörtern oder bei noch größeren Korpora noch aufwändiger zu beantworten wäre.

Church & Hanks (1989) waren die ersten, die zeigten, dass lexikalische Statistiken sinnvoll sind, um Konkordanzzeilen auf typische Verwendungsweisen hin zu analysieren. Sie verwenden dabei das statistische Maß der Mutual Information, das vereinfacht gesprochen, Worte in einer gewissen Nachbarschaft des Suchwortes (in der Regel zwischen 5 und 10 Wörtern oder auch innerhalb des ganzen Satzes) daraufhin bewertet, ob sie statistisch gesehen überproportional häufig mit dem Suchwort erscheinen. Eine Reihe von Datenbanken verwenden diese oder vergleichbare Kookkurrenzstatistiken, um die Gebräuchlichkeit von Wortverbindungen zu bewerten. Ein gutes Beispiel hierfür liefert das weit verbreitete Portal „Deutscher Wortschatz", welches zu einem Suchwort die statistisch signifikanten Kookurrenzen aufführt und darüber hinaus auch die signifikanten linken und rechten Nachbarn (s. Abb. 2)

**Signifikante Kookkurrenzen für Feindbild:**

ein (99), als (73), das (71), eins (68), neues (65), klares (54), Nummer (49), dient (45), Saudische (45), neoliberales (43), islamischen (42), irrsinnigen (42), auserkoren (41), Islam (40), : (40), geworden (38), Türk (35), Globetrotter (35), , (34), zum (33), Steuerstaates (32), Schmähgesänge (30), Juden (30), aufbaue (26), Westen (26), aufzubauen (25), gemeinsames (25), herhalten (25), dargestellt (24), willkommenes (24), Bayern-Fans (23), abgenutzt (23), löst (23), Jude (22), Bruchhagen (22), SPD-Arbeitsmarktexperte (21), Bill (21), Aug (20), Verhöhnung (20), aufgeheizten (20), Gates (19), Massen (19), Zelt (19), Terror (19), sein (19), zurecht (18), Sarkozy (17), Profilierung (17), Staatsmacht (17), entdeckt (17), beliebtes (16), ins (16), 68er (16), Amerika (16), Bushs (15), Brandner (15), Wanken (15), Israel (14), rückt (14), geliefert (14), USA (14), Husseins (14), entzündet (14), schlechthin (14), ist (14), Kuwait (13), Vernichtung (13), alte (12), Klientel (12), Mittleren (12), Kalten (12), parat (12), Gegners (12), größtes (12), vorm (12), Kapitalismus (11), Boll (11), Naturschutz (11), ausmachen (11), wilden (11)

**Signifikante linke Nachbarn von Feindbild:**

das (187), als (149), zum (149), neues (122), klares (95), ein (91), neoliberales (60), gemeinsames (48), alte (46), willkommenes (36), kein (35), sein (32), beliebtes (28), Das (28), gutes (26), größtes (24), wilden (23), großes (15), weiteres (15), ihr (12), " (11), dem (7), ins (6), vom (4)

**Signifikante rechte Nachbarn von Feindbild:**

Nummer eins (141), Nummer (113), herhalten (43), Islam (41), Bruchhagen (39), herhalten müssen (37), zurecht (36), geworden (36), entdeckt (35), auserkoren (30), USA (20), Israel (19), geliefert (19), geriet (19), dient (19), vieler (17), des (17), ab (17), " (15), Westen (14), Staat (13), gefunden (12), eines (10), der (9), vom (8), : (7), , (5), war (4)

Abb. 2: http://wortschatz.uni-leipzig.de (Abfrage vom 30.11.2010)

Zwar liefert eine rein statistische Methode einen sehr schnellen Überblick über die signifikanten Wortpaare innerhalb eines Satzes. Die Reihenfolge scheint je-

doch, was den syntaktischen Bezug zum Suchwort *Feindbild* angeht, zu einem gewissen Grade beliebig. So lässt sich zwar vermuten, dass die Artikel *ein*, *das* und die Adjektive *neues* und *klares* in Zeile (1) der Abbildung 2 einen klaren lokalen Bezug zum Suchwort haben, z.B. *ein Feindbild*, *neues Feindbild*, *klares Feindbild*. Schon weniges klar sind aber beispielsweise die Kookurrenzen in Zeile (3): *Türk*, *Globetrotter* oder *Schmähgesänge*. Die syntagmatischen Bezüge lassen sich hier nur erraten. Im Beispiel *Feindbild* ist diese syntaktische Unordnung noch durch etwas mehr Leseaufwand in den Griff zu kriegen, da es sich hier nur um etwa 50 Kookurrenzen handelt. Wirklich problematisch wird es jedoch bei Suchwörtern, die mehrere hundert oder sogar über 1000 Kookurrenzen haben. Dies ist kein so seltener Fall, wie in Abschnitt 4.2 erläutert wird. Man wird daher für die Beantwortung der oben gestellten Fragen nach typischen syntaktisch motivierten Wortverbindungen andere Verfahren benötigen, die auch syntaktische Informationen nutzen.

Ein für die syntagmatische Analyse weitergehender Ansatz stellt die von Adam Kilgarriff eingeführte Sketch Engine dar (Kilgarriff 2004). Die Sketch Engine geht über reine Kookkurrenzstatistiken insofern hinaus, als sie nur diejenigen Kookurrenzen berücksichtigt, die in einer vordefinierten syntaktischen Relation stehen. Solche Relationen können beispielsweise Adjektiv-Nomen, Verb-Objekt, Genitivattribute von Nomen oder Verb-Präpositionalphrasen Verbindungen sein. Sketch Engine Plattformen gibt es für so verschiedene Sprachen wie Englisch, Tschechisch, Japanisch oder Chinesisch. Eine einfache Übertragung des Sketch-Engine Ansatzes z.B. vom Englischen auf das Deutsche, ist aus wenigstens den beiden folgenden Gründen schwierig: die freie Wortstellung im Deutschen und der Kasussynkretismus führen dazu, dass eine Extraktion von syntaktischen Relationen auf der Basis von Wortarten und darauf basierenden Satzmustern, anders als im Englischen, zu keinen befriedigenden Ergebnissen führt. So haben Experimente mit der Sketch Engine für das Deutsche gezeigt, dass je nach Parametrisierung der Regeln, entweder die Analysegenauigkeit unzureichend ist oder aber die Abdeckung, d.h. der Anteil des analysierbaren Texts, zu gering ist (Kilgarriff 2004, Ivanova et al. 2008). Aus diesem Grund beruhen die beiden existierenden Ansätze für das Deutsche zur Extraktion von syntaktischen Relationen aus großen Textkorpora auf einem allgemeineren Formalismus, der syntaktische Satzfunktionen erkennen und lokale Mehrdeutigkeiten auflösen kann. Der erste, an der Universität Stuttgart entwickelte Ansatz zur Extraktion „signifikanter Wortpaare als Webservice“ (Fritzinger et al. 2009), beruht dabei auf dem Dependenzparser FSPAR (Schiehlen 2003), der zweite an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) entwickelte Prototyp, das DWDS-Wortprofil (Geyken et al. 2009), basiert auf Syncop

(Didakowski 2007), einem auf syntaktischem Tagging beruhenden Parsing-Formalismus.
In diesem Beitrag soll das DWDS-Wortprofil aus Nutzersicht erläutert werden. Dazu werden wir die Abfragefunktionalität, so wie sie sich für den Nutzer auf der Web-Plattform des DWDS präsentiert, am Beispiel des eingangs erwähnten Wortes *Feindbild* illustrieren (Abschnitt 4.1) und in Abschnitt 4.2 eine erste Einschätzung der Qualität des Wortprofils anhand des Vergleichs mit einem großen einsprachigen Wörterbuch gegeben werden. Abschnitt 5 gibt einen Ausblick auf die Potentiale des Wortprofils. Zuvor werden in Abschnitt 2 die Methode zur Erstellung von Wortprofilen und in Abschnitt 3 das derzeit verwendete DWDS-Wortprofil sowie dessen Datengrundlage beschrieben.

## 2. Das DWDS-Wortprofil

Das DWDS-Wortprofil ist das Ergebnis einer automatischen syntaktischen und statistischen Analyse sehr großer Korpora. Es liefert einen kompakten Überblick über die statistisch signifikanten syntagmatischen Beziehungen eines Wortes mit anderen Wörtern in Form einer Schlagwortwolke oder in Tabellenform. Die Berechnung des DWDS-Wortprofils erfolgt in drei Etappen, die in den folgenden Abschnitten näher beschrieben werden: Festlegung der zu extrahierenden syntaktischen Relationstypen, Extraktion der Relationen mittels einer automatischen syntaktischen Analyse und Bewertung der statistischen Signifikanz der extrahierten Relationen.

### 2.1. Festlegung der syntaktischen Relationen

Die Festlegung der syntaktischen Relationstypen erfolgt außerhalb der Berechnung der Relationen durch den Syntaxparser Syncop (s. Abschnitt 2.2), da die Relationstypen bestimmen, welche Analysen des Parsens in das Wortprofil einfließen. Die Liste der Relationstypen beruht auf den in Foth (2005) verwendeten Etiketten zur Markierung von Dependenzrelationen, die wiederum von SynCop verwendet werden. Typische im Wortprofil verwendete Etiketten sind ATTR oder OBJD. ATTR steht für eine Attribut-Relation zwischen Adjektiv und Nomen, OBJD für eine Dependenzrelation zwischen einem Verb und einer Nominalphrase im Dativ. Mit OBJD werden aus Gründen der Einfachheit für die Notation sowohl Dativobjekte als auch fakultative Ergänzungen oder freie Angaben zusammengefasst, da der Syncop über keine Verbsemantik verfügt und somit nicht entscheiden kann, ob es sich bei der dependenten Nominalphrase um ein Objekt oder nur eine fakultative Ergänzung oder Angabe handelt. Nicht alle für das Parsing verwendete Etiketten sind für

den Benutzer des Wortprofils relevant. Dies gilt beispielsweise für die Relation DET, die einen Artikel mit einem Nomen verbindet, die Relation APP, die festlegt ob eine Abfolge von Wörtern zu einer Nominalphrase gehören oder die Relation NEB, die das Verb eines Nebensatzes mit dem übergeordnete Wort verbindet. Diese Relationen sind zwar für den Parsing-Prozess wichtig, stellen aber für das Wortprofil nur Zwischenergebnisse dar.
Derzeit werden die folgenden Relationstypen für das DWDS-Wortprofil verwendet:
Relationstypen innerhalb von Phrasen

- Adjektiv-Nomen (Etikett: ATTR): *klares Feindbild, schöne Bescherung*
- Nomen-Koordination-Nomen (CJ): *grün und blau, Kopf oder Zahl*
- Nomen-Nomen (im Genitiv) (GMOD): *Abbau* des *Feindbildes, Abbau* von *Vorurteilen*
- Adverb-Adjektiv (ADVA): *sehr intelligent*, *hoch erfreut*

Phrasenübergreifende Relationen

- Nomen -Verb (SUBJ): das *Feindbild verblasst,* das *Badewasser läuft aus*
- Nomen-Verb (PSUBJ): die *Flasche* wurde *entkorkt*
- Nomen -Verb (OBJA): ein *Feindbild abbauen,* eine *Rede halten*
- Nomen -Verb (OBJD): dem *Publikum verkünden*
- Verb-Präposition-Nomen (V PP): *zur Verfügung stehen*, *auskommen ohne Feindbild*
- Verb-Verb (INFOBJ): *aufgehen sehen,* (auf etw.) *zu sprechen kommen*
- Adverb Verb (VADV)*: schallend lachen, freimütig zugeben*

## 2.2. Extraktion der syntaktischen Relationen

Wie bereits in Abschnitt 1 erwähnt, ist es für das Deutsche im Unterschied zum Englischen notwendig, die Extraktion syntaktischer Relationen auf der Basis eines syntaktisch voranalysierten Texts durchzuführen. Die syntaktische Analyse kann dabei aufgrund der speziellen Charakteristika des Deutschen (lange Abhängigkeiten und relative freie Wortstellung), nicht auf „flachen“ Sequenzabfolgen von Wortarten beruhen, sondern muss in der Lage sein, Dependenzinformationen für phrasenübergreifende Relationen aufzubauen. Ein solches Verfahren, welches sowohl syntaktische Satzfunktionen erkennen und lokale Mehrdeutigkeiten auflösen kann, ist das für das DWDS-Wortprofil verwendete Analysewerkzeug Syncop. Hier soll die Grundidee des Verfahrens dargestellt werden. Eine vollständige technische Beschreibung von Syncop findet sich in Didakowski (2007).
Das Parsing-Verfahren von Syncop beruht auf fünf Modulen, die sequenziell auf die Korpustexte angewandt werden:

(1) Satzsegmentierung und morphologische Analyse mit der TAGH-Morphologie (Geyken & Hanneforth 2005)
(2) Ermittlung der Phrasen und Annotation der darin enthaltenen Modizierer/Koordinationsfunktionen
(3) Ermittlung von dependenten bzw. koordinierenden Phrasen
(4) Ermittlung von Nebensätzen und Annotation der darin enthaltenen syntaktischen Köpfe
(5) Ermittlung von Hauptsätzen und Annotation der darin enthaltenen syntaktischen Köpfe

Bestandteile dieser Module sind außerdem:

- Die Auflösung von Kasus-Numerus-Genus Kongruenz. Dies ist für die Zuordnung von Attributivphrasen genauso wie für Subjekt-Verb Relationen von Bedeutung,
- Die Erkennung von Verb-Partikeln, die für die korrekte Lemmatisierung komplexer Verben in Verbzweitstellung notwendig ist,
- Präferenzregeln zur Auflösung mehrdeutiger globaler Satzanalysen,
- die Möglichkeit, syntaktische Regeln zu verletzen, um auch halb- oder ungrammatische Korpussätze zumindest partiell analysieren zu können.

Wie eingangs erwähnt, basieren Wortprofile auf sehr großen Korpora mit „real existierenden“, d.h. im Allgemeinen auch sehr komplexen oder unvollständigen Sätzen. Das Parsing-Verfahren zur Extraktion von Wortprofilen muss somit sowohl robust als auch effizient sein. Die Robustheit bedeutet in diesem Zusammenhang, dass das Verfahren nicht abbricht, wenn es bei einem komplexen Satz keine vollständige Satzanalyse liefern kann, sondern sich gegebenenfalls auch damit begnügt, partielle Analysen von Phrasen oder unterspezifizierte Satzfunktionen als Analyse zurückzugeben. Die Effizienz des Verfahrens wird dadurch gewährleistet, dass Syncop nicht beliebig eingebettete Sätze extrahiert – dies ist nur mit rekursiven Verfahren möglich –, sondern nur Sätze mit einer beschränkten Einbettung analysiert. Für praktische Zwecke der Korpusannotation ist dies aber durchaus ausreichend (Koskenniemi 1990).
Wie sich dies auf die Satzanalyse sowie die daraus zu extrahierenden syntaktischen Relationen auswirkt, soll an dem folgenden im DWDS-Kernkorpus enthaltenen Satz illustriert werden. Schwierigkeiten dieses Satzes sind die „langen“ Abhängigkeiten der passivischen Subjektfunktion (*Aspekt, beleuchtet*), die Passivierung des Gesamtsatzes sowie die komplexen Koordinationen (*Aspekt, Folgen, Auswirkungen*).

„Jeder Aspekt des Vertrags von Rom .. und alle Folgen und Auswirkungen, die ein britischer Beitritt nach sich ziehen dürfte, sind von allen Seiten beleuchtet worden.“ Hervorhebung (A.G.) (Archiv der Gegenwart 36, 1966)

Von Syncop wird dieser Satz folgendermaßen analysiert (die Analyse beschränkt sich dabei auf die für die Extraktion wichtigen Satz- und Koordinationsrelationen).

[[Jeder*@DetN* Aspekt*@HEAD*]$_{np}$*@SUBJ* [des*@DN* Vertrags*@HEAD*]$_{np}$*@GN* … und*@CC* [alle*@DN* Folgen*@HEAD* und*@CC* Auswirkungen*@HEAD*]$_{np}$*@SUBJ* sind*@FAUXV* … beleuchtet*@FMAINV* worden*@FAUXV* .]$_{cl_passive}$

In dieser Klammernotation werden syntaktische Funktionen (Kopf- und Modifiziererrelationen) durch das Symbol @ markiert. Im Beispiel steht dabei @HEAD für die Kopffunktion, @Subj für die Subjektrelation, @CC für die Koordination, die anderen Funktionen stehen für Modifiziererrelationen: *@DN* für eine Nomen-Determinierer Relation, *@GN* für die Nomen-Genitiv Relation, *@FMAINV* für das Hauptverb, *@FAUXV* für die Verb-Auxiliar-Relation. Die Markierung der Sätze und Phrasen werden mit eckigen Klammern notiert: Im Beispiel gibt es einen Satz, der als Passiv-Klausel markiert (cl_passive) ist und drei Phrasen, die jeweils als Nominalphrase (NP) annotiert werden. Zur einfacheren Lesbarkeit ist die Klammernotation in Abbildung 4 in der Dependenzansicht dargestellt.




Abb. 3: Dependenzbaum

Aus der Satzanalyse des Beispielsatzes durch den Syncop-Parser lassen sich die folgenden syntaktischen Relationen (s. 2.1) für das Wortprofil extrahieren, wie sich dem Dependenzbaum ablesen lässt:

- PSUBJ: (Aspekt , beleuchten)
- GMOD: (Aspekt, Vertrag)

- CJ: (Aspekt, Folge)
- CJ: (Aspekt, Auswirkungen)
- PSUBJ: (Folge, beleuchten)
- PSUBJ: (Auswirkung, beleuchten)

Zusammenfassend lässt sich somit festhalten, dass das Analysesystem Syncop in der Lage ist, Phrasen, Hauptsätze (Klauseln) und Nebensätze (Sub-Klauseln) eines Satzes zu markieren sowie die dazugehörigen syntaktischen Funktionen (Kopf- und Modifiziererrelationen) zu ermitteln. Dabei gilt folgende Einschränkung: Bei komplexen oder unvollständigen Sätzen versucht Syncop nicht, eine vollständige Syntaxanalyse des Satzes zu erzeugen, sondern gibt auch partielle Analysen zurück. Aus den Analysen wiederum werden die syntaktischen Relationen extrahiert.

### 2.3. Berechnung der statistischen Salienz

Die extrahierten syntaktischen Relationen werden mittels eines statistischen Verfahrens bezüglich ihrer statistischen Signifikanz gewichtet. Dafür werden die aus den Texten extrahierten Relationen aufsummiert und danach bewertet, ob sie überproportional häufig vorkommen. Wichtig hierbei ist, dass für die Zählung weder einfache Wortformen noch Lemmaformen, sondern Lemma-Wortart-Paare verwendet werden. Ein Lemma-Wortart-Paar ist beispielsweise (*grau*, Adjektiv) wie in *graue Eminzenz* oder (*grau*, Modifizierer) wie in *grau meliert*. Der Grund, weshalb man Lemma-Wortart-Paare und nicht einfach Lemmata verwendet, besteht darin, dass Lemmata bezüglich ihrer Wortart mehrdeutig sein können und somit andere Relationstypen besitzen. Beispielsweise ist (*grau*, Adjektiv) Attribut eines Nomens, wohingegen (*grau*, Modifizierer) Attribut eines Adjektivs ist. Diese Vermischungen umgeht man, indem man für die Lemma, Wortart – Paare unterschiedliche Relationstabellen aufbaut.

Das im DWDS-Wortprofil verwendete Maß der Salienz (Lin 1998) ist eine Erweiterung der Mutual Information (MI). Bei der Salienz wird wie für die MI berechnet, ob eine Wortkombination im Korpus „häufiger als erwartet" vorkommt. Allerdings bezieht die Salienz hierbei nicht die relativen Häufigkeiten über das gesamte Korpus mit ein, sondern schränkt diese auf den spezifischen syntaktischen Relationstyp ein, über den die Wortkombination in Verbindung steht. Formal ausgedrückt wird die Salienz eines Tripels (w, r,w') wie folgt definiert:

$$Sal\ (w,r,w') - \frac{\|w,r,w'\| \times \|*,r,*\|}{\|w,r,*\| \times \|*,r,w'\|} \quad (1)$$

In der Formel (1) sind w und w' Lemma-Wortart-Paare, r ein syntaktischer Relationstyp. $\|w,r,w'\|$ steht für die Anzahl der Vorkommen des Tripels $(w,r,w')$ im annotierten Korpus. $'*'$ ist ein Platzhalter für eine beliebiges Lemma-Wortart-Paar, und $\|w,r,*\|$ ist definiert als die Summe der Häufigkeiten über alle Lemmata $wj'$ mit $\|w,r,w_j'\|$. Entsprechend wird $\|*,r,*\|$ definiert als die Summe über alle Tripel $(w,r,w_j')$, die miteinander über die Relation $r$ in Verbindung stehen. Diese Formel entspricht der Formel, die Church & Hanks (1989) für die MI vorgeschlagen haben mit dem zusätzlichen Faktor $\|*,r,*\|$. In Übereinstimmung mit Kilgarriff (2004) lässt sich feststellen, dass das Maß der Salienz im Vergleich zum Maß der MI den Vorteil hat, Tripel mit geringer Häufigkeit nicht überzubewerten.

## 3. Datengrundlage und Parametrisierung des DWDS-Wortprofils

Die Korpora, die als Datengrundlage des DWDS-Wortprofils dienen, sind grundsätzlich frei wählbar. Die Zusammensetzung und Größe der Korpora spielen für das Wortprofil jedoch eine wichtige Rolle.

Die Zusammensetzung der Korpora ist insofern relevant, als die extrahierten syntaktischen Relationen die im Korpus vorkommenden syntaktischen Nachbarn des Wortes widerspiegeln. Daher erhöht ein breit gestreutes, nach Textsorten ausgewogenes Korpus, ein sogenanntes allgemeinsprachliches Referenzkorpus, die Qualität des Wortprofils in Bezug auf die allgemeinsprachliche Aussagekraft. Spezialkorpora oder spezielle Zeitungskorpora werden somit andere Wortprofile liefern als Referenzkorpora.

Auch die Korpusgröße hat einen großen Einfluss auf die Wortprofile, denn Wortprofile sind in der Regel nur aussagekräftig, wenn das Lemma wenigstens 500, besser jedoch 1000 Mal im Korpus auftaucht (siehe auch Ivanova et al. 2008). Unter dieser Zahl ist die Aussagekraft eines Wortprofils nur begrenzt, da viele syntaktische Relationen dann in der Regel nur ein oder zwei Mal vorkommen und somit kaum nachweisbar ist, dass es sich bei den extrahierten syntaktischen Relationen um typische Beispiele und nicht um Zufallsfunde handelt. Insofern spielt auch die absolute Korpusgröße eine Rolle, als sich mit wachsender Korpusgröße auch die Anzahl der verschiedenen Wörter erhöht, die hochfrequent im Korpus vorkommen. Dabei stellt sich heraus, dass eine Korpusgröße von 100 Millionen laufenden Textwörtern zu klein ist, um eine für die zu erwartende Benutzungssituation ausreichende Anzahl von Wortprofilen zu extrahieren. So gibt es beispielsweise im 100 Millionen Textwörter umfassenden DWDS-Kernkorpus (s. unten) nur etwa 5.000 Lemmata, die mehr

als 1.000 Mal vorkommen. Daraus ergibt sich, dass man weitaus größere Korpora benötigt, wenn man qualitativ ausreichende Wortprofile für eine Stichwortanzahl erstellen möchte, die zumindest der eines Kompaktwörterbuchs entspricht.
Für das Deutsche liegen bislang keine Referenzkorpora geschriebener deutscher Standardsprache mit einer nach Textsorten ausgewogenen Verteilung vor[1]. Bei der Erstellung eines hinreichend großen Wortprofils ist man somit entweder auf Spezialkorpora oder auf eine Mischung aus Referenz und Zeitungskorpora angewiesen. Der derzeitige Prototyp des DWDS-Wortprofils (fortan kurz: Wortprofil) beruht auf einer Mischung eines Referenz- und eines Zeitungskorpus mit einer Gesamtgröße von 500 Millionen Tokens: dem DWDS-Kernkorpus, einem nach Textsorten ausgewogenen und zeitlich gleichmäßig gestreuten Referenzkorpus der deutschen Sprache des 20./21. Jahrhunderts (Geyken 2007, www.dwds.de/textbasis), ca. 110 Millionen Tokens) und dem ZEIT-Archiv (1946–2009, ca. 400 Millionen Tokens, www.zeit.de). Daraus wurden für etwa 90.000 Lemmata Wortprofile extrahiert mit etwa 2.000.000 syntaktischen Relationen. Knapp ein Fünftel aller Wortprofile basiert auf Lemmata, die mindestens 1000 Mal im Korpus vorkommen, also gemäß des o.g. Schwellwerts ausreichend für ein umfassendes Wortprofil sind. Als Schwellwert für die Mindestanzahl des Vorkommens einer syntaktischen Relation wurde die Zahl 4 gewählt. Damit soll verhindert werden, dass okkasionelle Verbindungen fälschlicherweise in das Wortprofil aufgenommen werden. Ob der Schwellwert von 4 zu hoch ist, lässt sich nur experimentell und in Abhängigkeit von einer Fragestellung beantworten. In der Folge soll die Nutzung sowie die Qualität des gegenwärtigen Wortprofils anhand einiger Beispiele illustriert werden.

## 4. Beispiele einiger Wortprofile

### 4.1. Feindbild

Das DWDS-Wortprofil kann über die Webseite des lexikalischen Informationssystems DWDS (www.dwds.de) genutzt werden. Im DWDS-Informationssystem werden für ein Suchwort, welches der Benutzer in die Suchmaske eingibt, die Ergebnisse in mehreren Fenstern (panels) präsentiert. Es gibt drei Typen von Fenstern: Wörterbuch-, Korpus- und Statistikfenster. Die Fenster lassen sich frei zu sogenannten Sichten (views) zusammenstellen. Das DWDS-

1 Die Situation ist anders, wenn man Webcorpora betrachtet, also Korpora, die mit Crawling-Techniken aus dem Web gewonnen wurden. Dafür gibt es für das Deutsche bereits ein 1,6 Milliarden Token großes Korpus (Pomikalek et al. 2009). Allerdings bleibt zu zeigen, dass dessen Aussagekraft vergleichbar mit einem Korpus ist, dessen Quellen sich aus gesicherten, redaktionell bearbeiteten Texten speist.

Wortprofil ist eines der Fenster und Teil der Standardsicht des DWDS-Informationssystems.
Anhand des bereits in Abschnitt 1 erwähnten Wortes *Feindbild* sollen die verschiedenen Recherchemöglichkeiten mit dem Wortprofil im DWDS-Informationssystem illustriert werden. Die Standardansicht des Wortprofils erfolgt in Form einer Schlagwortwolke (Abb. 4). Man sieht, wie bei Schlagwortwolken üblich, signifikante Begriffe in einer größeren Schrift als weniger signifikante. Dabei basiert die Signifikanz auf der in Abschnitt 2.3 erläuterten Salienz: je salienter eine syntaktische Relation, desto größer die Schriftart. Im Unterschied zu einer klassischen Schlagwortwolke werden hier somit die Begriffe über die reine statistische Signifikanz auch nach der syntaktischen Relevanz gewichtet (s. Abschnitt 1).
Für das bereits weiter oben erwähnte Wort *Feindbild* werden im DWDS-Wortprofil 32 verschiedene syntaktische Relationen mit insgesamt 384 Vorkommen extrahiert. Die Voraussetzung für die Aufnahme einer Relation in das Wortprofil ist, dass dafür wenigstens vier Belege im Korpus vorkommen. Abgedruckt ist die im Korpus am häufigsten verwendete flektierte Form.




Abb. 4: Wortprofil in Schlagwortwolkenansicht

Neben der Darstellung in Form der Schlagwortwolke kann das Wortprofil auch in der klassischen Tabellenansicht präsentiert werden (Abb. 5).

DWDS-Wortprofil

relevanteste syntaktische Relationen für Feindbild als **Substantiv** Frequenz: 384 — Relationenfilter

| Stammform | Wortart | syntaktische Relation | Salience | Frequenz |
|---|---|---|---|---|
| Vorurteil | Substantiv | Feindbild hat die Beiordnung Vorurteile | 14.65 | 15 |
| stimmen | Verb | stimmt hat Aktiv-Subjekt Feindbild | 11.42 | 18 |
| klar | Adjektiv | klares ist Attribut von Feindbild | 10.66 | 39 |
| Abbau | Substantiv | Abbau hat Genitivattribut Feindbild | 10.39 | 9 |
| antibolschewistisch | Adjektiv | antibolschewistische ist Attribut von Feindbild | 9.02 | 4 |
| alt | Adjektiv | alten ist Attribut von Feindbild | 7.71 | 103 |
| aufbauen | Verb | aufbaut hat Akkusativobjekt Feindbild | 7.09 | 8 |
| taugen_als | PP | taugt hat Vergleichskonjunktion Feindbild | 6.7 | 4 |
| verblassen | Verb | verblaßt hat Aktiv-Subjekt Feindbild | 6.37 | 4 |
| auskommen_ohne | PP | auskommt hat Präpositionalphrase ohne_Feindbild | 5.74 | 4 |

Abb. 5: Wortprofil in Tabellenansicht

In der Tabellenform werden in der ersten Spalte alle statistisch signifikanten Wortformen aufgeführt (Spalte Stammform). Spalte 2 verzeichnet die Wortart, bei mehreren Wörtern auch den Phrasentyp (beispielsweise PP, s. letzte Zeile der Abbildung). In Spalte 3 ist der syntaktischen Relationstyp beschrieben. Die Wortformen in Spalte 1 sind nach absteigender Salienz (s. Spalte 4) geordnet. Durch Klick auf die Spaltenüberschrift „Frequenz“ (Spalte 5) wäre auch eine Sortierung nach der Frequenz möglich.
Die syntaktisch relevanten Nachbarn von *Feindbild* sind in den folgenden syntaktischen Relationstypen zu finden:

- Adjektiv-Nomen (Etikett: ATTR): *altes, antibolschewistisches, gemeinsames, gepflegtes, ideales, ideologisches, intaktes, klares, klassisches, linkes, neues, primitives, richtiges, schlichtes, verblasstes, westliches, äußeres, überkommenes Feindbild*
- Nomen-Nomen (im Genitiv) (GMOD): *Abbau*, *Verlust*
- Nomen-Koordination-Nomen (CJ): *Vorurteil*
- Nomen-Verb (SUBJ): *bleiben*, *stimmen*, *verblassen*
- Nomen-Verb (OBJA): *abbauen*, *aufbauen*, *brauchen*, *nehmen*, *schaffen*
- Verb-Präposition-Nomen (V PP): *auskommen_ohne*, *taugen_als*

Die Relationstypen lassen sich über das Wortprofil-Fenster ansteuern, indem man den Relationenfilter anklickt (Abb. 4). Es werden dann die syntaktischen Relationstypen aufgeklappt (Abb. 6). Klickt man dann auf eine der Relationstypen, beispielsweise auf „Attribut“, erhält man alle Wortformen, die in einer Attributrelation (Adjektiv-Nomen) zum Wort *Feindbild* stehen. Diese Filter können bei hochfrequenten Wortprofilen sehr nützlich sein. Beispielsweise hat das geläufige Substantiv *Haar* 13509 Relationen (davon 532 verschiedene) im DWDS-Wortprofil. Eine Darstellung als Schlagwortwolke wäre hier sehr unübersichtlich. Durch das Filtern nach einzelnen Relationstypen hingegen erhält man homogene Listen und handhabbare Größen.




Abb. 6: Filtern nach Relationstyp Attribut

Ein wesentlicher Mehrwert des Wortprofils besteht darin, dass alle extrahierten Relationen stets mit den dazugehörigen Satzkontexten im Korpus verknüpft sind und somit einen Überblick über den Verwendungszeitraum und die semantischen und pragmatischen Kontexte ermöglichen, in denen die syntaktische Relation verwendet wird. Klickt mal beispielsweise in Abbildung 5 auf das erste Wort *Abbau*, welches zum Suchwort *Feindbild* in einer Attributrelation im Genitiv steht (Relation GMOD), gelangt man zu den insgesamt neun Kontexten, die vom Analysesystem Syncop aus dem Wortprofil-Korpus extrahiert wurden (Tabelle 1). Die Beispiele sind nummeriert und nach absteigendem Datum geordnet: der jüngste Beleg stammt aus dem Jahre 2000, der älteste von 1982. Die möglichen Textsorten (Belletristik, Gebrauchstext, Wissenschaft, Zeitung) sind in der dritten Spalte aufgeführt; diese ist mit einem Doppelpunkt von der Belegquelle getrennt. In der 4. Spalte findet sich der Kontext für die syntaktische Relation, wobei die gesuchten Wörter farblich hervorgehoben sind.

1 2000-06-29 Zeitung: ZEIT Wir brauchen einen Abbau der Feindbilder von den „illegalen Einwanderern“, bei denen in der Regel Täter und Opfer verwechselt werden.

2 1998-04-23 Zeitung: ZEIT Wir brauchen einerseits einen Abbau des Feindbilds von den illegalen Einwanderern, bei denen ohnehin in der Regel Täter und Opfer verwechselt werden.

3 1990-04-20 Zeitung: ZEIT Und was den Abbau der Feindbilder betrifft, die ja schon vor 1945 von Deutschland gegen die Sowjetunion und von der UdSSR gegen Deutschland geschaffen wurden:

4 1990-02-23 Zeitung: ZEIT Der Abbau der Feindbilder sollte den Sowjetbürgern eine europäische Identität und der Perestrojka westlichen Elan bringen.

5 1989-05-26 Zeitung: ZEIT Er mahnte aber auch die Bundesrepublik, in der Abrüstung nicht hinter der DDR zurückzubleiben (die gerade Truppen einseitig reduzierte) und wünschte sich einen Abbau des Feindbildes auch in der Bundesrepublik.

6 1989-05-05 Zeitung: ZEIT Bobowikows Beitrag ließ auch über die Wahlen hinaus unverblümt die ideologische Kriegserklärung gegen die Grundwerte der Perestrojka erkennen – gegen Glasnost im Inneren und gegen den Abbau der Feindbilder.

7 1988-10-21 Zeitung: ZEIT Sie verfolgt jedoch nicht mehr die Illusion des Ausspielens und Aufspaltens, weil sie das Ansehen und die Integration der Sowjetunion in die Weltgemeinschaft durch den Abbau der ideologischen Feindbilder fördern will.

8 1988-09-30 Zeitung: ZEIT Mit dem Abbau der alten Feindbilder tut sich die DDR offenbar schwer.

9 1982-05-15 Zeitung: ADG Er fordere deshalb die Demonstranten auf, auf beiden Seiten für einen Abbau der Feindbilder einzutreten, die Resignation gegenüber der weltweiten Aufrüstung abzuwehren und eine aktive gewaltfreie persönliche Haltung anzustreben.

Tab. 1: Konkordanzansicht für Relation GMOD: Abbau, Feindbild

Die Stärke des syntaktischen Ansatzes zeigt sich bei der Extraktion von syntaktischen Funktionen. Hierfür reicht in der Regel nicht nur ein lokaler Kontext, sondern der gesamte Satz muss ausgewertet werden, da Komplemente von Verben im allgemeinen Fall nicht mit lokalen Funktionen extrahiert werden können. Ein Beispiel hierfür liefert die Relation Verb-Präpositionalphrase *auskommen ohne Feindbild*. Diese Relation enthält alle Sätze im Korpus, bei denen das Verb *auskommen* – mit und ohne abtrennbarem Verbpartikel – in Relation mit der Präpositionalphrase *ohne Feindbild* steht. Durch Klick auf *auskommen ohne* in Abbildung 4 gelangt man zu den Satzkontexten in Tabelle 2,

bei denen die Konstruktion sowohl in Verbzweit- wie auch Verbletztstellung vorkommt.

**auskommen_ohne:**

1 1993-10-22 Zeitung: ZEIT Hamburg 1993; 206 S., 26, -DM fast ganz ohne dieses Feindbild auskommt:

2 1991-09-26 Zeitung: ZEIT Hysterie und Haß. Doch die Unfähigkeit des einstigen Freiheitshelden gegen Kreml und Kommunismus, der jetzt ohne Feindbilder nicht mehr auskommt, der über den Weltmarkt nichts, aber über die Weltverschwörung gegen Georgien alles weiß, hat Hysterie und Haß gesät.

3 1989-10-20 Zeitung: ZEIT „Das Land kam vorübergehend ohne Feindbilder aus „, heißt es im Begleitbuch.

4 1972-07-28 Zeitung: ZEIT Das kritische Denken des Autorenteams kommt ohne Feindbild nicht aus.

Tab. 2: Konkordanzansicht für Relation V-PP: auskommen, ohne Feindbild

## 4.2. grau

Im vorangegangen Abschnitt wurden die verschiedenen Nutzungsmöglichkeiten des Wortprofils auf der Webplattform www.dwds.de erläutert. Noch nicht angeschnitten wurde die lexikographische Qualität des Wortprofils. Aufgrund des vom Korpus abgedeckten Zeitraums ist hierfür der Vergleich mit einem großen einsprachigen deutschen Gegenwartswörterbuch naheliegend. Für den Vergleich haben wir zwei große Wörterbücher herangezogen: das große Wörterbuch der deutschen Sprache in 10 Bänden des Dudenverlags (GWDS) und das Wörterbuch der deutschen Gegenwartssprache (WDG). Da der Vergleich relativ ausführlich dargestellt werden soll, um die Bandbreite der möglichen Vergleichsparameter zu illustrieren, soll er an dieser Stelle nur für ein Wort demonstriert werden, nämlich für das Adjektiv *grau*. Dieses Adjektiv haben wir ausgewählt, weil es häufig genug für ein ausgeprägtes Wortprofil ist, weil es mehrere Lesarten hat und weil es keine grundsätzlichen Bedeutungsveränderungen in den letzten Jahren erfahren hat. Aus urheberrechtlichen Gründen beziehen wir uns nur auf den Artikel des WDG (siehe Anhang), weisen aber an den entsprechenden Stellen auf die Unterschiede zum GWDS hin. In der Tat zeigt sind die Artikelbeschreibungen des Stichworts *grau* in beiden Wörterbüchern in ihrer Bedeutungsstruktur im großen und ganzen vergleichbar, wenn auch nicht gleich. Einige wichtige Unterschiede zwischen beiden Wörterbüchern lassen sich feststellen. Das GWDS hat die Lesarten 1 und 2 des WDG, bei der der konkrete und der bildliche Gebrauch beschrieben wird, zu einer zusammengefasst. Das GWDS führt im Gegenzug eine neue Lesart ein,

bei der *grau* im Sinne von „sich an der Grenze der Legalität bewegen" (umgangssprachlich) verwendet wird, wie beispielsweise in *graue Händler*. Darüber hinaus verzeichnet das GWDS einige geläufige syntaktische Relationen, die das WDG nicht aufführt, beispielsweise *graue Schläfen*, *graues Brot* oder *grau meliert*. Umgekehrt fallen auch einige Relationen weg, wie beispielsweise *graue Theorie*, *graue Maus* oder *in Ehren ergraut*. Vom Umfang her halten sich beide Wörterbücher die Waage.
Der quantitative Vergleich der Wörterbücher mit dem Wortprofil zeigt zunächst einmal folgendes: das WDG verzeichnet 43 syntaktische Relationen, das GWDS nur 25 Relationen. Viele der Relationen im WDG sind jedoch „erwartbar", wie beispielsweise *Bart* oder *Haarsträhne*, wenn man ohnehin *Haar* aufführt. Insofern bietet das WDG hier keinen Bedeutungsüberschuss gegenüber dem GWDS. Im DWDS-Wortprofil werden 7727 Relationen extrahiert, darunter 388 verschiedene Relationen (f>=4, sal>0). Dies ist mehr als das Zehnfache der im Wörterbuch aufgeführten Wortverbindungen. Es stellen sich somit die Fragen nach der gemeinsamen Schnittmenge von Wortprofil und WDG (markiert als (+WP;+WDG)), nach denjenigen Kookurrenzen, die im Wortprofil, aber nicht im WDG sind (+WP;-WDG), und umgekehrt nach der Qualität derjenigen Wortverbindungen, die nicht im Wortprofil, dafür aber im WDG sind (-WP;+WDG).

(+WP;+WDG)

Zunächst einmal ergibt der Abgleich von Wortprofil und WDG, dass sich in der Schnittmenge folgende 29 Kookurrenzen mit *grau* befinden:
*Vorzeit, Maus, Haare, Alltag, Bart, Himmel, Star, Wolke, Mauer, Theorie, Altertum, Haarsträhnen, Uniform, Augen, Haupt, Kostüm, Gestein, alt, färben, Scheitel, Elend, Öde, Regenwolken, Wölkchen, Morgen, Meer, Gesicht, Haut, Strand*
Nimmt man noch Synonyme oder nahe Hyperonyme hinzu, so findet man auch weitere zwei WDG-Einträge im WP: Frühlicht (WP: *Morgenlicht*), Ferne (WP: *Unendlichkeit*)

(+WP;-WDG)

Über 90% aller salienten syntaktischen Relationen, genauer 369 der 398 Relationen des Wortprofils, sind nicht im WDG. Diese Zahl sagt zunächst nichts über die lexikalische Relevanz dieser Relationen aus. Bei unserem Vergleich haben wir daher die Einträge auf ihre lexikalische Relevanz angesehen, wobei wir einen Unterscheidung zwischen hoch salient (sal>10), salient (sal>5) und schwach salient (sal<5) gemacht haben.

Hoch salient sind insgesamt 54 syntaktische Relationen, 10 davon im WDG; unter den verbleibenden finden sich folgende lexikographisch relevante Wortverbindungen (geordnet nach absteigender Salienz):

- *graue Eminenz*
- *graue Maus* (fehlt im WDG im übertragenen Sinne)
- *grauer Kapitalmarkt*
- *graue Schläfen*
- *grauer Pfandbriefmarkt* (im Sinne von: geschlossener Fonds)
- *Graue Ackerschnecke* (limax agrestis)
- *graue Zellen*

Ferner fehlen folgende Eigennamen-Kontexte:

- *Graue Panther* (politische Partei)
- *Graue Wölfe* (extremistische politische Partei)
- *Graue Kloster* (Gymnasium in Berlin)

Salient (Sal>5) sind weitere 143 Wortverbindungen im Wortprofil; davon sind 12 im WDG. Unter den verbleibenden 132 Relationen sind die lexikographisch relevanten Wortverbindungen (geordnet nach absteigender Salienz):

- *graues Einerlei*
- *graues Umweltpapier*
- *graue Asche* (von Toten oder von verbranntem Material)
- *graue Habenzinsen* (auf dem nicht freien Kapitalmarkt)
- *grauer Markt*
- *grauer Haarkranz*
- *grauer Nadelstreifen*
- ferner alle Adjektiv-Modifizierer: beispielsweise 'schütter' wie in *schütteres graues Haar*, ‚voll' wie in *voll grauer Haare*

Schwach salient: unter den Wortverbindungen mit einer Salienz < 5 sind nur noch wenige lexikographisch interessante Verbindungen zu finden; u.a. *graue Erbsen*, *graue Liste* oder der *Graue Burgunder*. Die meisten sind transparent und erwartbar, wie beispielsweise *graue Krawatte*, *graue Socke* oder *graue Matte*.

(-WP;+WDG)

Umgekehrt finden sich 14 Wortverbindungen im WDG, die nicht im Wortprofil auftauchen:

*Frühlicht, Stoff, wie eine Maus, bei Nacht sind alle Katzen grau, keine grauen Haare wachsen lassen, in Ehren grau geworden, Gesichtsfarbe, Woche, Tag, Monat, Mittelalter, vor grauen Jahren, Ferne, Zukunft*
Diese Zahl scheint auf den ersten Blick recht hoch, relativiert sich aber schnell, wenn man sie anhand ihrer Gebräuchlichkeit analysiert und mit den Alternativen vergleicht, die das Wortprofil für einige dieser Wortverbindungen bereitstellt. In der Folge soll dies anhand der Lesarten des WDG detaillierter gezeigt werden.

## WP und WDG im Lesartenvergleich

Das WDG unterscheidet insgesamt fünf Lesarten von *grau*.

**Lesart 1**: /Mischfarbe aus Schwarz und Weiß/
(+WP;+WDG): *Auge, Gestein, Haut, Himmel, Kostüm, Mauer, Maus, Meer, Regenwolke, Strand, Uniform, Wolke*
(+WP;-WDG): unter anderem *Granit, Beton, Anzug, Kittel, Kutte*
(-WP;+WDG): *Stoff*. Dafür sind im Gegenzug im WP einige Synonyme, wie *Flanell*, *Samt* oder *Wolle*. Die Vergleichskonjunktion *grau wie eine Maus* findet sich nur im WDG (auch nicht im GWDS); im Gegenzug ist aber im WP das wesentlich häufiger gebrauchte *graue Maus* enthalten, mit beispielsweise folgendem Beleg: *Arminia Bielefeld hat schon seit ein paar Jahren den Ruf als „graue Maus" der Liga weg*. Die ZEIT, 26.12.2008

**Lesart 2a:** /farblos, bleich, durch zunehmendes Alter/
(+WP;+WDG): *Bart, Haar, Haarsträhne, Haupt, Scheitel, Star, alt*
(+WP;-WDG): u.a. *Schläfen, Strähne, Haarkranz*
(-WP;+WDG):
*sich keine grauen Haare wachsen lassen* (bildlich, umgangssprachlich im Sinne von sich keine Sorgen machen); wird als Konstruktionsmuster vom Wortprofil derzeit nicht extrahiert, ist aber unter den Beispielen von „grau Haar" mehrfach zu finden.
*in Ehren grau geworden*: taucht in den Korpora insgesamt 4 Mal auf, ist aber im Wortprofil nicht salient; aber dafür gibt es über 100 Treffer im Korpus für "in Ehren ergraut". Diese sind unter dem Wortprofil von „Ehre" zu finden.

**Lesart 2b**: /farblos, bleich, durch Blutleere/
(+WP;+WDG): *Gesicht*
(+WP;-WDG): keine Einträge
(-WP;+WDG):

*graue Gesichtsfarbe*: ist im WP nicht enthalten, da es bereits im zugrundeliegenden Korpus mit insgesamt nur drei Belegen unter dem derzeit gewählten Salienzschwellwert von 4 liegt.
*vor Haß* (kommt im WDG nur als Literaturbeleg vor (Marchwitza)). In der Korpora gibt es diesen Beleg gar nicht. Auch im Web findet man hierfür nur einen Treffer[2] – und dieser stammt noch dazu aus dem WDG: „ Sie starrte mich grau vor Haß an“ (Marchwitza Jugend 248)

**Lesart 3:** /übertragen/, trostlos und öde
(+WP;+WDG): *Alltag, Elend, Elend, Morgen, Theorie, öde*
Letztere Konstruktion ist im WDG nur durch das Goethezitat belegt: *grau, treuer Freund, ist alle Theorie* . Im Wortprofil finden sich dazu 140 Belege mit anderen Konstruktionen, die illustrieren, dass „graue Theorie“ nicht nur im Zusammenhang mit dem Goethe-Zitat gebraucht wird, z.B.

- Das Ethos ist kein Istzustand, es ist ein Sollzustand, trotzdem nicht graue Theorie, sondern eine sehr praktische Sache. (ZEIT, 04.01.2010)
- Und jede Eröffnungsvorbereitung graue Theorie sein lässt. (ZEIT, 28.8.2009)
- Dass dies keine graue Theorie ist, zeigt ein Fall in den USA. (ZEIT, 18.5.2009)
- Alles graue Theorie, die uns Winzer in die Illegalität treibt. (ZEIT, 21.11.2008)
- Doch wie sich zeigte, blieb das graue Theorie. (ZEIT, 27.8.2008)
- Die Betriebshandelsspanne erweist sich, nämlich, als graue Theorie. (ZEIT, 24.7.2008)
- …

(+WP;-WDG): keine Einträge
(-WP;+WDG): *Monat*, *Tag* und *Woche*. Alle drei Begriffe entstammen einem Literaturbeleg von Feuchtwanger (s. Anhang). Diese Kookurrenzen sind nicht salient, dafür sind aber nahe Synonyme im Wortprofil: *Herbsttag*, *Nachmittag*, *Novembertag* oder *Vormittag.*
*das graue Elend kriegen* (sich tief unglücklich fühlen, zeigen) ist nicht im Wortprofil, dafür aber die einfachere Attribut-Nomen-Relation *graues Elend.*

**Lesart 4**: /weit zurückliegend längst vergangen/
(+WP;+WDG): *Altertum, Vorzeit*
(-WP;+WDG): *Mittelalter*

2 Suchanfrage unter www.google.de vom 6.12.2010

**Lesart 5**: /unbestimmt, ungewiss/
(+WP;+WDG): keine Relationen sind im Schnitt
(-WP;+WDG): *graue Ferne, Zukunft*; diese sind beide nicht im Wortprofil salient.

Zusammengefasst noch einmal die wichtigsten Schlüsse, die sich aus dem Beispiel *grau* ableiten lassen. Wortprofile können im Vergleich zu großen einsprachigen Wörterbüchern ein Vielfaches der syntaktischen Relationen zu einem Wort enthalten. So verzeichnet das Wortprofil zu *grau* knapp 400 verschiedene Relationen, wohingegen das WDG 43, das GWDS nur 25 Wortverbindungen aufführen. Dies schlägt sich im auch direkten Vergleich nieder: mit einer hohen Salienz (sal>5) enthält das Wortprofil mehr als 20 lexikographisch relevante Beispiele, die nicht im WDG verzeichnet sind. Bei einer Salienz von unter 5 im Wortprofil nimmt die Dichte der lexikographisch relevanten Relationen stark ab. Von den knapp 200 syntaktischen Relationen (s<5) sind nicht mehr als 5 als lexikographisch relevant einzustufen. Erstaunlich ist zunächst, dass das Wortprofil nur etwa 70% der im Wörterbuch verzeichneten Wortverbindungen als syntaktische Relation enthält. Zu den meisten dieser fehlenden Verbindungen führt das Wortprofil jedoch gebräuchlichere Alternativen auf; in anderen Fällen existiert die Wortverbindung nur als Literaturzitat. Man findet im Wortprofil auch eine zusätzliche Lesart, die zwar nicht im WDG, jedoch im GWDS verzeichnet ist. Das Wortprofil hat aber hier die gebräuchlicheren Beispiele: *grauer Kapitalmarkt* oder *grauer Markt* (WP) statt *graue Händler* oder *graues Material* (GWDS). Schließlich, das zeigt das Beispiel *graue Theorie*, findet man mit dem Wortprofil zahlreiche authentische Kontexte und Verwendungen, die von dem einzigen dazu im WDG aufgeführten Goethezitat (*grau, mein Freund, ist alle Theorie*) abweichen. Eine Konstruktion übrigens, die in das GWDS nicht mehr aufgenommen wurde.

## 5. Potentiale des Wortprofils

Das Beispiel *grau* in Abschnitt 4.2 macht deutlich, dass in dem gegenwärtigen DWDS-Wortprofil mit einem zugrundeliegenden Korpus der Größe von 500 Millionen Textwörtern eine große Reichhaltigkeit von Wortverbindungen steckt, die es auch im Vergleich mit großen einsprachigen Wörterbüchern interessant macht. Bevor wir zum Abschluss dieses Beitrags kurz auf die Potenziale dieses Ansatzes eingehen, sollen zuvor noch ein paar einschränkende Anmerkungen zur Reichweite der in 4.2 gemachten Beobachtungen gemacht werden.

Erstens wird es notwendig sein, diese Befunde auf der Basis aller offenen Wortarten anhand einer hinreichend großen Anzahl von Wörtern zu replizieren. Eine entsprechende Studie, die einen Vergleich von 50 Wortprofilen mit entsprechenden Einträgen in WDG-Artikeln ähnlich dem in Abschnitt 4.2 durchführt, ist derzeit im Gange. Dabei zeigt sich, und das ist die zweite Relativierung, dass die Abdeckung der Relationsextraktion mit dem Parser Syncop für den Verbbereich im Allgemeinen weitaus geringer ist als für den nominalen und den adjektivischen Bereich. Dies ist deshalb unmittelbar einleuchtend, da viele syntaktische Relationstypen von Nomina und Adjektiven mit lokalen Abhängigkeiten zu erfassen sind, wohingegen bei den Verben syntaktische Funktionen, also phrasenübergreifende Abhängigkeiten zu extrahieren sind.
Die dritte Relativierung schließlich betrifft die Evaluierung der Qualität der extrahierten Relationen in Bezug auf deren Status als Kollokation. Hierzu lässt sich folgendes feststellen: Syntaktische Relationen weisen, wenn sie statistisch signifikant sind, eine Ähnlichkeit zu dem weit verbreiteten Kollokationsbegriff von Hausmann (1984) auf, weichen aber in zwei Aspekten entscheidend davon ab. Kollokationen im Hausmannschen Sinne bestehen aus einer Kollokationsbasis und einem Kollokator. Beispiele hierfür sind *schütteres* **Haar, Termin** *einhalten* oder *hoch* **erfreut** (die Basis ist hier fettgedruckt, der Kollokator kursiv). Syntaktische Relationstypen hingegen sind symmetrisch, d.h. wenn das Wort A in Relation mit dem Wort B steht, so steht auch umgekehrt B in Relation mit A. Der Grund hierfür ist ein pragmatischer, da bei automatisch extrahierten syntaktischen Relationen Basis und Kollokator nicht sicher bestimmt werden können und man somit die Abdeckung des Wortprofil-Ansatzes für den Nutzer erhöht, wenn man die Umkehrbarkeit der Relation und somit den Zugriff auf die Kollokation sowohl über die Basis als auch den Kollokator zulässt. Ganz abgesehen davon ist es bei einer Datenbankrepräsentation immer möglich, die Kollokation sowohl unter der Basis als auch unter dem Kollokator unterzubringen. Darüber hinaus sind bei Kollokationen im Hausmannschen Sinne die Menge der Kollokatoren für eine gegebene Basis sehr eingeschränkt. Vor allem geht es darum, „nicht erwartbare" Kollokatoren zu beschreiben. In einem völlig automatisierten System ist dies nicht möglich. Hier lassen sich für ein gegebenes Wort die syntaktisch relevanten Kookurrenzpartner nur über die statistische Signifikanz ermitteln. Diese aber ist im allgemeinen nicht gleichbedeutend mit einer Erwartbarkeit. Beispielsweise führt das DWDS-Wortprofil für das o.g. Nomen *Haar* 223 Adjektive auf, die allesamt statistisch signifikant in Bezug auf das Suchwort sind. Darunter gibt es eine ganze Reihe von „echten", weil nicht erwartbaren Kollokationen, wie beispielsweise *schütter*, *gegelt* oder *gescheitelt*, aber auch andere erwartbare signi-

fikante Kollokate wie *schwarz*, *blond*, *lang* oder *nass*. Obzwar diese Kookurrenzen keine Kollokationen im Hausmannschen Sinne sind, so sind viele davon auch nicht beliebig, sondern semantisch oder pragmatisch motiviert. Beispiele hierfür wären für *Ball* beispielsweise *Ball abspielen*, *Ball zuwerfen*, oder *Recht anwenden*, *Recht brechen*, *Recht auf Mitbestimmung* für *Recht*.

Jenseits dieser drei Einschränkungen sind jedoch die Potenziale künftiger Entwicklungen von statistischen Wortprofilen enorm. Die Erweiterungen liegen zunächst in der Verbesserung der Abdeckung der Parser zur Extraktion der syntaktischen Relationen, vor allem aber in der Erweiterung und Vergrößerung der Korpusbasis. Wie bereits weiter oben erwähnt, lassen sich mit größeren, nach Textsorten breit gestreuten Korpora sinnvolle Wortprofile für weitaus mehr Stichwörter bilden und die Konfidenz der Salienz syntaktischer Relationen verbessern.

Darüber hinaus sind Wortprofile als Ressource für die Textproduktion von großen Wert. Mit statistischen Wortprofilen lassen sich sowohl typische Wortverbindungen für ein Wort auffinden als auch deren semantische und pragmatische Kontexte in authentischen Textzusammenhängen mit wenigen Mausklicks nachschlagen. Hierzu noch einmal ein Beispiel: Im GWDS beispielsweise findet man, wenn man unter dem Stichwort *grau* die Relation *grau meliert* nachschlägt, einen authentischen Textbeleg:

grau melierte Schläfen; sein Haar war inzwischen g. meliert, was seine Erscheinung noch würdevoller machte (Danella, Hotel 30) (GWDS, Eintrag grau, Lesart 1)

Das Beispiel deutet an, dass sich die Relation *grau meliert* in der Regel auf eine Person bzw. dessen Haar bezieht. Die Schlussfolgerung darüber, ob sich dies nur auf eine männliche Person bezieht oder ob Träger desselben positiv konnotiert wird, bleibt jedoch der Intuition des Nutzers überlassen. Mit dem Wortprofil lässt sich diese Vermutung mit einem Klick an nicht nur einem, sondern vielen weiteren authentischen Korpusbeispielen überprüfen. So findet man im DWDS-Wortprofil die Verbindung *grau meliert* mit einer Salienz von 38 und einer Frequenz von 51, die sich ausschließlich auf männliche Personen beziehen. Darüber hinaus belegen zahlreiche dieser Kontexte die positiven Konnotationen, die entweder auf den sozialen Status des Trägers oder andere pragmatische Kontexte zurückzuführen sind. Jeweils ein Beispiel hierzu:

Eduardo Montes ist ein stolzer Mann. *Grau meliertes* Haar, hellblaues Maßhemd, fein gepunktete Krawatte – fast zu elegant für sein profanes Chefzimmer in der Hofmannstraße. (ZEIT, 04.10.2006)

Oder dass man gute Kunden zum Essen in einen der sechs Speisesäle einlädt, wo *grau melierte* Kellner auf Kosten des Hauses schon mal einen Grand Cru entkorken. (ZEIT, 31.10.2001)
Künftige Experimente mit anderen und größeren Korpora werden zeigen, inwieweit Werkzeuge dieser Art ebenso natürlich für das Verfassen von Texten sein werden, wie heutzutage bereits die Rechtschreib- und Grammatikkontrolle oder die Synonymvorschläge in Textverarbeitungen.

## Literaturverzeichnis

Braun, S.; Kohn, K. und J. Mukherjee (2006). *Corpus Technology and Language Pedagogy*. Peter Lang. Frankfurt.

Church, K. W., und Hanks, P. (1989). Word association, norms, mutual information, and lexicography. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 76–83.

Didakowski, J. (2007): SynCoP - Combining syntactic tagging with chunking using WFSTs. Linguistik in Potsdam, In: Proceedings of FSMNLP 2007. Universitätsverlag Potsdam.

Fritzinger, F., Kisselew, M., Heid, U., Madsack, A., Schmid, H. (2009). Werkzeuge zur Extraktion von signifikantenWortpaaren als Web Service. Vortrag: GSCL Symposium Sprachtechnologie und eHumanities, Duisburg, 26-27 Februar 2009

Foth, K. (2005). Eine umfassende Dependenz-Grammatik des Deutschen. Universität Hamburg.

Hausmann, F.-J. (1984). Wortschatzlernen ist Kollokationslernen. In: Praxis des neusprachlichen Unterrichts. 31. Jg. (1984), S. 385-406.

Geyken, A.; Hanneforth, Th. (2005). TAGH: A Complete Morphology for German based on Weighted Finite State Automata. In: Proceedings of FSMNLP 2005, Lecture Notes in Artificial Intelligence. Springer.

Geyken, A. (2007): *The DWDS corpus: A reference corpus for the German language of the 20th century*. In: Fellbaum, Christiane (ed.): Collocations and Idioms: Linguistic, lexicographic, and computational aspects. London: Continuum Press, S. 23-41.

Geyken A., Didakowski, J. und Siebert, A. (2009). Generation of word profiles for large German corpora. In Yuji Kawaguchi, Makoto Minegishi and Jacques Durand (ed.). Corpus Analysis and Variation in Linguistics, p. 141–157.

Ivanova, K., Heid, U., Schulte im Walde, S., Kilgarriff, A, Pomikálek, J. (2008). Evaluating a German Sketch Grammar: A Case Study on Noun Phrase Case. Proceedings of the 6th Conference on Language Resources and Evaluation. Marrakech, Morocco.

Jurish, B., (2003). A Hybrid Approach to Part-of-Speech Tagging. Final report, Project Kollokationen im Wörterbuch, BBAW, Berlin.

Kilgarriff, A.; Rychly, P., Smrz, P., and D.Tugwell (2004). The Sketch Engine. In Proceedings Euralex 2004. Lorient, France, July: 105-116.

Koskenniemi, K. (1990). Finite State Parsing and Disambiguation. In Proceedings of COLING, vol. 2, 229-232.

Lin, D. (1998). Automatic Retrieval and Clustering of Similar Words. COLING-ACL98. Montreal.

Pomikalek, J.; Rychly, P. and Kilgarriff, A (2009). *Scaling to Billion-plus Word Corpora*. Advances in Computational Linguistics. Special Issue of Research in Computing Science Vol 41, Mexico City.

Schiehlen, M. (*2003). A cascaded finite-state parser for. German*. In Proceedings of the 10th EACL, Budapest.

## Wörterbücher

[GWDS] Duden - Das große Wörterbuch der deutschen Sprache in 10 Bänden (1999). Bibliographisches Institut, Mannheim. 3. Auflage.

[WDG] Klappenbach, R. und Steinitz, W. (Hg.) (1964-1977). *Wörterbuch der deutschen Gegenwartssprache (WDG)*. Berlin: Akademie-Verlag.

## Anhang

### Wörterbuchartikel *grau* im WDG

**grau** /Adj./ **1.** /*Mischfarbe aus Schwarz und Weiß*/ die g. Haut des Elefanten; g. wie eine Maus; ein g. Stoff, Kostüm; eine g. Uniform; er hat g. Augen; g. Gestein; g. Mauern; Am grauen Strand, am grauen Meer /Und seitab liegt die Stadt STORM 8,194; g. Rauch-, Regenwolken; der Himmel ist ganz g.; g. Frühlicht; Grau ist"s immer, wenn ein Morgen naht BRECHT *Gedichte* 278; /*sprichw.*/ bei Nacht sind alle Katzen g. (*nachts erkennt man keine äußerlichen Unterschiede*) ein helles, dunkles, düsteres, bleiches, kaltes, fahles, farbloses Grau; das Grau der Regenwolken; das Blau des Himmels ging in Weiß und Grau über; eine Bluse in Grau; die Dame in Grau; sie kam in Grau, bevorzugt die Farbe Grau; *vgl.* Grauchen **2.** *farblos, bleich* **a)** *durch zunehmendes Alter*: g. Haar; eine g. Haarsträhne; Sie hatte bereits einen grauen Scheitel G. HAUPTM. 4,552; ein g. Bart; Achtung vor einem g. Haupte (*einem alten Menschen*) haben sein g. Haar färben; alt und g. werden; er ist in Erhren g. geworden; der g. Star (*krankhafte Trübung der Linse im Auge*) /*bildl.*/ u m g . darüber, deshalb brauchst du dir keine g. Haare wachsen zu lassen (*darüber brauchst du dir keine Sorgen zu machen*) **b)** *durch Blutleere*: ein g. Gesicht; mein mageres graues Gesicht und die Trostlosigkeit meines Blickes BÖLL *Wort* 12; eine g.

Gesichtsfarbe; Der alte Mann wurde ganz grau im Gesicht Brecht *Dreigroschenroman* 296; Sie starrte mich grau vor Haß an Marchwitza *Jugend* 248 **3.** */übertr./ trostlos, trübe, öde*: Es hieß warten, einen grauen Morgen und einen grauen Tag und eine graue Woche und einen grauen Monat Feuchtw. *Tag* 78; In Dresden ging der graue Alltag wieder los Renn *Kindheit* 23; Grau, teurer Freund, ist alle Theorie Goethe *Faust* I 2038; ihr erschien die Welt g. und öde; u m g . s c h e r z h . das g. Elend kriegen (*sich tief unglücklich fühlen, zeigen*)◎ ich kann das Grau in Grau unserer Nachkriegsepoche auf die Dauer nicht aushalten G. Hauptm. *Sonnenuntergang* I **4.** *weit zurückliegend, längst vergangen*: in g. Vorzeit; im g. Altertum, Mittelalter; vor g. Jahren, Zeiten; **5.** *unbestimmt, ungewiß*: das liegt noch in g. Ferne, Zukunft;

Quelle: WDG – elektronische Version, www.dwds.de

# Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen[1]

**Cyril Belica – Institut für Deutsche Sprache Mannheim**

Abstract

Dieser Beitrag betrachtet lexikalisch-semantische Relationen aus einer emergentistischen Perspektive vor dem Hintergrund eines korpusgeleiteten empirisch-linguistischen Ansatzes. Er skizziert, wie eine systematische Erfassung und Auswertung des Kookkurrenzverhaltens[2] von Lexemen – die Analyse der Ähnlichkeit von Kookkurrenzprofilen mit Hilfe von selbstorganisierenden lexikalischen Merkmalskarten und ihre im Diskurs verankerte Interpretation – wichtige Einblicke in die Struktur verschiedenartiger Verwendungsaspekte dieser Lexeme einschließlich ihrer semantischen Nähe ermöglichen. Die vorgestellte Methodik wird dabei – über die explorativ-analytischen Zielsetzungen hinaus – als eine abduktive, auf Theoriebildung zielende Generalisierungsstrategie im postulierten Lexikon-Syntax-Kontinuum verstanden. Zum Schluss werden die Anwendungsmöglichkeiten einiger Komponenten dieser Methodik in der Lexikografie, Lexikologie und Didaktik diskutiert.

## 1. Das Deutsche Referenzkorpus

In Anlehnung an den Auftrag des IDS, „[…] die deutsche Sprache in ihrem gegenwärtigen Gebrauch zu dokumentieren und zu erforschen […]“, hat der Aufbau von elektronischen Korpora am Institut eine lange Tradition. Der in den 60er Jahren des zwanzigsten Jahrhunderts noch neue Gedanke, Analyse und Beschreibung der deutschen Sprache auf computergespeicherte Textkorpora zu stützen, wurde von den Gründungsdirektoren des IDS Paul Grebe und Ulrich Engel in die Tat umgesetzt: Im Jahr 1964 begannen die konzep-

1 Dieser Beitrag bietet einen Rückblick auf die Forschungsarbeit des Programmbereichs Korpuslinguistik am Institut für Deutsche Sprache (IDS) in Mannheim in den letzten Jahren mit einem besonderen Fokus auf die Hypothese der emergenten Natur lexikalisch-semantischer Relationen. Er ist durch eine stilistische Bearbeitung des Transkripts eines am 28. März 2008 in Brixen gehaltenen Vortrags entstanden und bewahrt dessen kursorischen Charakter. Er stellt die Ergebnisse gemeinsamer Arbeit des Autors mit Holger Keibel, Marc Kupietz und Rainer Perkuhn und der Kooperation mit Marie Vachková vor, denen der Autor auch für die Durchsicht und viele Verbesserungen des Manuskripts herzlich dankt.

2 Wegen der klassifikatorischen Vorbelastung des Terminus *Kollokation* in der deutschsprachigen Fachliteratur wird am IDS der Bezeichnung *Kookkurrenz* insbesondere dann der Vorzug gegeben, wenn es darum geht, die theorieneutrale, prä-interpretative *Perspektive* auf die im lokalen lexikalischen Kontext empirischen Sprachmaterials objektiv beobachtbaren (d.h. messbaren) Kohäsionsstrukturen hervorzuheben.

tuellen Arbeiten und die Texterfassung (damals noch auf Lochkarten und -streifen) des *Mannheimer Korpus 1*.
Heute gehört das Deutsche Referenzkorpus des IDS mit seinen ca. 3,4 Milliarden laufenden Textwörtern (Stand: November 2008) weltweit zu den größten und wichtigsten sprachwissenschaftlich motivierten Sammlungen des geschriebenen Deutsch. Es wird im Hinblick auf Größe, Stratifikation, extratextuelle Dokumentation und urheberrechtliche Unbedenklichkeit fortlaufend weiterentwickelt. Wegen der restriktiven Rechtslage ist DeReKo kein offenes Korpusarchiv. Das IDS erwirbt das nicht übertragbare und durch weitere Auflagen eingeschränkte Nutzungsrecht an DeReKo-Textmaterial durch Lizenzverträge mit den (derzeit mehr als 170 einzelnen) Textgebern und räumt seinerseits den DeReKo-Endbenutzern das Recht zur wissenschaftlichen, nichtkommerziellen Nutzung von DeReKo durch individuelle Nutzungsvereinbarungen ein, wie Kupietz und Keibel (2009a) berichten. Sie setzen fort:

To retain its long-time outstanding reputation as a trustworthy partner of text donors in Germany and abroad, the IDS maintains the highest standards of ethical and meticulous legal conduct pertaining to the DeReKo use and encourages and expects its employees to report any suspected violations of the end user licence agreement.

Vor dem Hintergrund kritischer Überlegungen zum wissenschaftstheoretischen Status empirischen Materials in der linguistischen Forschung (vgl. Keibel / Kupietz, 2009) wird DeReKo seit Anfang der 1990er Jahre nicht als ein „ausgewogenes“ oder gar „repräsentatives“ Korpus, sondern als eine möglichst umfassende *Ur-Stichprobe* (*primordial sample*) des öffentlichen Schriftsprachgebrauchs konzipiert. Die Komposition von geeigneten auf der DeReKo-Ur-Stichprobe basierenden *virtuellen Korpora* ist bei dieser Strategie aus der Phase des Korpus*aufbaus* herausgelöst und bleibt – unter Einbeziehung der spezifischen Zusammenhänge der konkreten linguistischen Fragestellung und ihrer Implikationen bezüglich der Adäquatheit der zugrundezulegenden Grundgesamtheit – den späteren Korpus*nutzungs*phasen vorbehalten. So wurde zum Beispiel das der aktuellen Version der Kookkurrenzdatenbank *CCDB* (s. Abschnitt 4.1) zugrundeliegende virtuelle Korpus *ccdb-2007* u.a. durch eine per Zufallsziehung vorgenommene quantitative Dämpfung der thematischen Komponenten „Politik“ und „Sport“ aus DeReKo abgeleitet.

## 2. Der theoretische Rahmen

Die prinzipielle Nützlichkeit von Sprachkorpora in der linguistischen Forschung und in den benachbarten Anwendungsgebieten ist heutzutage weitestgehend unumstritten. Allerdings wird die Rolle des empirischen Materials in Abhängigkeit von der jeweiligen Zielsetzung, dem gewählten theoretischen Standpunkt, von verschiedenen praxisbezogenen Faktoren und nicht zuletzt vom aufgebrachten Maß an wissenschaftsmethodischer Strenge sehr unterschiedlich verstanden. Der „dokumentarisch-unterstützende" und der „Vermutungen überprüfende" Beweggrund sind wohl die zwei vorherrschenden Motivationen eines typischen DeReKo-Nutzers. Im sprachtheoretischen Kontext wird das Korpus darüber hinaus oft nur als ein – aus der Perspektive einer frei wählbaren Systematik – möglichst erschöpfend *zu beschreibender* Datenkörper gesehen.

Will man jedoch einen korpus*geleiteten* empirisch-linguistischen Ansatz konzipieren, der *explanatorische Theoriebildung* zum Ziel hat, so ist es ratsam, zunächst das einzusetzende Gerüst von Annahmen und Lösungsstrategien darauf zu hinterfragen, ob es mit dem epistemischen Status des Forschungsgegenstandes (d.h. der Sprache) überhaupt vereinbar ist. In ihrer Betrachtung eines genuin korpuslinguistischen Forschungsprogramms, in dessen Kontext auch die im vorliegenden Aufsatz beschriebenen Arbeiten entstanden sind, stellen Kupietz und Keibel (2009b) ihre Auffassung der weit verbreiteten – obgleich ungesicherten – Annahmen und Lösungsstrategien der linguistischen Theoriebildung folgendermaßen dar:

> Vermeiden sollte man zunächst vor allem die folgenden Credos der theoretischen Linguistik – nicht weil sie erwiesenermaßen falsch wären, sondern eher, weil nicht erwiesen ist, dass sie zutreffen: erstens die Annahme, dass Sprache als formales System adäquat fassbar ist; zweitens die Annahme, dass Dekomposition uneingeschränkt als Explanationsprinzip anwendbar ist; drittens, eine vollständige Theorie als Forschungsziel anzustreben; und viertens, die sprachliche Kompetenz zum Gegenstand der Forschung zu machen.

Sie diskutieren unter anderem, wie man bei dem Versuch, die „breite Kluft zwischen der angestrebten theoretischen Beschreibungsebene einerseits und der phänomenologisch zugänglichen Ebene des Sprachgebrauchs andererseits" zu schließen, überhaupt zu hinreichend abgesicherten Erkenntnissen für eine explanatorische Theoriebildung kommen kann und betonen in diesem Zusammenhang „die Notwendigkeit, sich dem Untersuchungsgegenstand Sprache mit möglichst wenigen Vorannahmen über diesen Gegenstand selbst zu nähern". Im Rahmen eines derart konzipierten Forschungsprogramms wird

das Phänomen *Sprache* aus einer emergentistischen Perspektive ausgeleuchtet, der zufolge „alles Regelhafte und Konventionelle in der Sprache ein Epiphänomen des Sprachgebrauchs ist und von den Sprachteilnehmern fortlaufend ausgehandelt wird". Dieses ontologische Prinzip – so die Grundhypothese – durchdringt die Sprache auf allen ihren Ebenen und ist z.B. an der Konventionalisierung syntaktischer Strukturen und an der – besonders im Hinblick auf das Hauptthema dieses Beitrags relevanten – Entstehung und Festigung denotativer und konnotativer Aspekte lexikalisch-semantischer Relationen konstitutiv beteiligt. Seine Wirkung – so die psychologische Arbeitshypothese, die hier nur angedeutet werden soll – manifestiert sich in der Interaktion von zwei miteinander verschränkten kognitiven Mechanismen[3]:

- Im spontan-assoziativen Kontext nimmt der sehr weit gefasste Begriff der *Ähnlichkeit* eine Schlüsselrolle ein. Unsere Fähigkeit, im Alltag auch sehr unkonkrete, verschwommene Ähnlichkeiten unwillkürlich und schnell zu erkennen – besser gesagt, unsere Unfähigkeit, das permanente unwillkürliche Erkennen von Ähnlichkeiten außer Kraft zu setzen – ist lebensnotwendig, da Situationen und Kontexte, in denen wir uns bemühen, uns zurechtzufinden, nie vollkommen identisch sind. Ohne die spontane Sensibilität für Ähnlichkeiten wären frühere Erfahrungen für zukünftiges Verhalten nicht ohne analytisches, schlussfolgerndes Denken nutzbar, sie wären nicht assoziativ generalisierbar. Das gilt insbesondere auch für sprachliche Generalisierungen.
- Im symbolisch-klassifikatorischen Kontext werden gezielt regelhafte Zusammenhänge in der individuellen Erfahrungswelt aufgespürt, typischerweise in Form von abstrakten A-posteriori-Generalisierungen. Im Bereich der Sprache manifestieren sich solche ordnenden Mechanismen schon in der Fähigkeit von Sprachteilnehmern, auf der Basis ihrer Spracherfahrung Ad-hoc-Hypothesen über grammatische Regelhaftigkeiten zu formulieren. Bei Ontogenese, Phylogenese und bei jedem einzelnen Sprachverarbeitungsakt eilen die spontan-assoziativen kognitiven Strategien des „Sich-Zurechtfindens" den symbolisch-klassifikatorischen generell zeitlich voraus.

Eine dementsprechend zentrale Rolle kommt in dem hier vorgestellten methodologischen Ansatz dem Begriff der Ähnlichkeit – „als einem graduellen Gegenkonzept zur *Identität*" – zu (Kupietz / Keibel 2009b: 44). Es wird angenommen, dass man – von normativen bzw. terminologischen Zusammenhängen abgesehen – den vermöge gradueller Ähnlichkeit zwischen Situationen

3 Vgl. die Annahme „[…] human language faculty is instantiated in the mind as an equilibrium between two interdependent general cognitive agencies […]" bei Vachková / Belica (2009).

und Kontexten entstandenen unscharfen Konzeptualisierungen von sprachlichen Phänomenen zwar mit Hilfe von A-posteriori-Klassifikationen eine Pseudo-Identität verleihen kann (um sie symbolisch referenzierbar zu machen), dass aber diesen Explikationen immer eine Verzerrung innewohnt, z.B. in Form von Vereinfachung oder Überzeichnung.[4]
Wenn aber Regelhaftes und Konventionelles wie angenommen im Sprachgebrauch (und gemessen am Erfolg der Kommunikation) ausgehandelt wird, dann ist es auch plausibel anzunehmen, dass die im Diskurs fortlaufend stattfindenden Aushandlungsprozesse konkrete Spuren in Sprachkorpora – d.h. in den Aufzeichnungen von Kommunikationsprozessen im Allgemeinen – hinterlassen, welche mit geeigneten strukturentdeckenden Methoden zu ihren Entstehungsbedingungen zurückverfolgt werden können. Im Hinblick auf die hier diskutierten lexikalisch-semantischen Relationen bedeutet es, dass man die Rekonstruktion dieser Aushandlungsprozesse mit einer „Spurensuche" in zwei verschiedenen „Strata" gleichzeitig vorantreiben sollte: in den *lokalen* lexikalischen Kontexten, in denen die einzelnen lexikalischen Einheiten gebraucht wurden einerseits, und in den *globalen*, situativen Kontexten, in denen jene Aushandlungsprozesse jeweils stattgefunden haben andererseits. Da die zu betrachtenden – lokalen und globalen – Kontexte untereinander im Allgemeinen nie vollkommen identisch sind (s.o.), wird man, um die Überprüfung ihrer Zugehörigkeit zu einem bestimmten Aushandlungsakt und somit um einen Hinweis für ihre fallbezogene Relevanz bestrebt, die Explikation ihrer Identität möglichst meiden bzw. hinauszögern und stattdessen wie oben dargelegt mit dem graduellen Konzept der Ähnlichkeit arbeiten müssen.
Im Folgenden wird gezeigt, wie man diese allgemeinen Überlegungen im Einzelnen mit Hilfe einer großen Anzahl von Kookkurrenzprofilen, die auf Kookkurrenzanalysen höherer Ordnungen basieren, und unter Anwendung des Funktionsprinzips von selbstorganisierenden Karten (Kohonen 1990), einer Art von künstlichen neuronalen Netzen, schrittweise operationalisieren und zu *lexikalischen Merkmalskarten* (*Lexical Feature Maps*) gelangen kann, die u.a. wertvolle Einsichten in die semantische Struktur lexikalischer Einheiten bieten. Ein wesentlicher Unterschied dieser Herangehensweise zu anderen, überwiegend auf *information retrieval* ausgerichteten Ansätzen, wie etwa der *Latent Semantic Analysis* (*LSA*, s. zum Beispiel Deerwester et al. 1990), besteht darin, dass hier die den multidimensionalen semantischen Raum definierenden Einheiten ihrerseits durch komplexe Kookkurrenzstrukturen definiert sind, in denen die in den *lokalen* Kontexten auftretenden Regularitäten und Fluktuatio-

4 Selbstverständlich sind derartige Explikationen ungeachtet aller damit verbundenen Schwierigkeiten oft unentbehrlich und sinnvoll.

nen detaillierter festgehalten werden können als bei der dokument-orientierten Verfahrensweise. Der zweite wichtige Unterschied liegt in der postulierten psychologischen Verankerung dieser Kookkurrenzstrukturen begründet, eine Annahme, die insbesondere bei der Auseinandersetzung mit syntagmatischen Mustern (s. Abschnitt 3.1) als recht naheliegend erscheint: Auf diese Art definierte syntagmatische Muster weisen eine derart hohe Rekurrenzrate in den uns vorliegenden Aufzeichnungen der Kommunikationsprozesse auf, dass es – wenn man einmal das Konzept der graduellen Ähnlichkeit als das konstituierende Prinzip von sprachlichen Generalisierungen akzeptiert hat (s. oben im Abschnitt zu *spontan*-assoziativen kognitiven Mechanismen) – geradezu unverantwortlich erscheint ohne einen triftigen Grund die Vermutung zu verwerfen, dass sie von individuellen Sprachteilnehmern – und dennoch intersubjektiv – mitunter als sprachliche *Gestalt* wahrgenommen und bei Sprachproduktion und -rezeption auch als Ganzes verarbeitet werden. Sie können also insofern als *psychologisch reale* Objekte betrachtet werden, als sie möglicherweise beim Sprachverstehen nicht zwingend immer dekomponiert werden, wiewohl sie im Sinne traditioneller sprachsystemischer Kategorien dekomponierbar sind (und analog bei der Sprachproduktion). Nähere Aufschlüsse dazu und eine auf experimentelle Daten gestützte Argumentation werden von den geplanten interdisziplinären Untersuchungen im Schnittpunkt von Korpuslinguistik, Psycholinguistik und kognitiver Psychologie erwartet.

## 3. Lokaler Kontext

### 3.1. Kookkurrenzanalyse

Zur Erfassung von Regularitäten in den lokalen lexikalischen Kontexten in sehr großen Korpora wird am IDS meistens die Analysemethode „Statistische Kollokationsanalyse und -clustering“ (Belica 1995) eingesetzt. Sie implementiert einen erweiterten iterativen Algorithmus zur Extraktion von *Kookkurrenzen höherer Ordnung*, welche als diskontinuierliche n-Tupel (im Unterschied zu *n-Grammen*) von Lexemen mit variierenden relativen Positionen zueinander (anders als *positionsgebundene n-Gramme*) auftreten können. Es werden standardmäßig Kookkurrenzen beliebiger Ordnung ermittelt, wobei das auszuwertende Kontext-Fenster dynamisch bestimmt wird. Die extrahierten Kookkurrenzen werden mit Hilfe eines Clusteringverfahrens in Form von Baumhierarchien angeordnet, welche die zunehmend feinkörnige Kookkurrenzstruktur einzelner Ordnungen visualisieren. Das Analysemodul ist seit 1995 in das IDS-Korpusrecherchesystem COSMAS (*Corpus Search,*

*Management and Analysis System)* integriert und online aufrufbar. Abbildung 1 zeigt einen kurzen Ausschnitt aus einer Web-basierten Präsentation der Ergebnisse der Kookkurrenzanalyse für das Wort *machen*. Die in der letzten Spalte aufgelisteten syntagmatischen Muster (z.B. „macht sich|mir … große Sorgen") können als stellvertretende Illustrationen der eigentlichen Kookkurrenzen verstanden werden und helfen hauptsächlich, die aufgedeckten Kookkurrenzcluster einerseits zurück auf Sprachdaten, andererseits auf die Intuition eines kompetenten Sprechers zu beziehen. Sie sind oft ein nützlicher Ausgangspunkt für die qualitative Interpretation einzelner Kookkurrenzen (vgl. Perkuhn 2007).

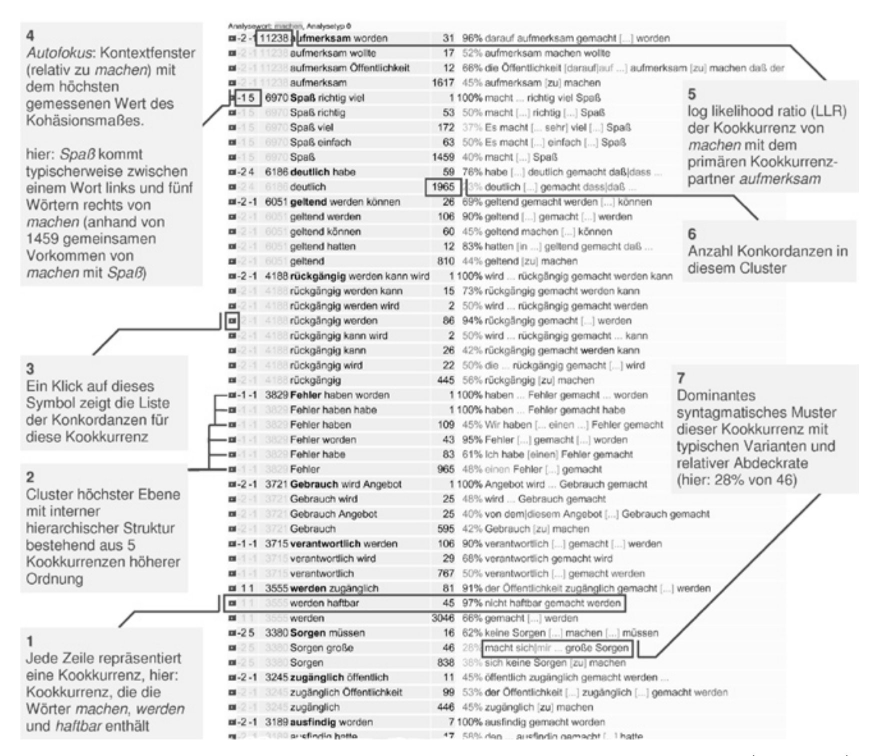



Abb. 1: Annotierte Web-Präsentation der Kookkurrenzanalyse des Wortes machen (Ausschnitt)

Das Analysemodul ist sprachunabhängig, wie im Anhang A anhand der Analyse des Wortes *why* in einem englischsprachigen Korpus illustriert wird.

### 3.2. Kookkurrenzprofile

Die Gesamtheit aller quantitativen Ergebnisse der Kookkurrenzanalyse zu einem gegebenen Analyseobjekt (einem Lexem, einer Wortverbindung usw.) wird als *Kookkurrenzprofil* des Objektes bezeichnet und stellt – informell gesagt – ein Kondensat seines Gebrauchs dar. Es erfasst sowohl dominante Wortverbindungsstrukturen wie auch subtile Varianzphänomene im lokalen lexikalischen Kontext des analysierten Objektes, und bietet dadurch eine detaillierte Auskunft über die syntagmatische und paradigmatische Einbettung des Objekts im Sprachgebrauch aus präferenzrelationaler[5] Sicht.

## 4. Globaler Kontext

Die Argumentation zu situativen Kontexten im Abschnitt 2 ließ die Frage offen, wie diese Kontexte, in denen konventionalisierte lexikalisch-semantische Relationen in der Interaktion der Sprachteilnehmer ausgehandelt werden, erfasst, bzw. in den Korpora aufgespürt werden können. Im Folgenden wird gezeigt, dass diese globalen Kontexte sich in der Struktur wechselseitiger Beziehungen einer großen Anzahl von Kookkurrenzprofilen manifestieren, obwohl dies auf den ersten Blick überraschen mag, da Kookkurrenzprofilen nur die lokalen Kontexte zugrunde liegen.

### 4.1. *CCDB* – das empirisch-linguistische Methodenlabor

Um die wechselseitigen Beziehungen zwischen Kookkurrenzprofilen eingehend untersuchen zu können, wurde im Jahr 2001 am IDS eine „korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs“ (s. Keibel / Belica 2007) begründet. In ihrem Kern enthält sie große Volumina sekundärer empirischer Daten, eine Kookkurrenzdatenbank, die durch mehrfache, unterschiedlich parametrisierte Kookkurrenzanalysen des gesamten Grundformenvokabulars eines großen Korpus[6] errechnet wurde. Derzeit enthält sie mehr als 220.000 nicht-triviale Kookkurrenzprofile pro Parametersatz, bis zu 250 Top-Level-Clusterknoten pro Kookkurrenzprofil und bis zu 100.000 Konkordanzzeilen pro Kookkurrenzcluster. Da diese Analysen mit Hilfe des im Abschnitt 3.1 beschriebenen Analysemoduls durchgeführt wurden, kann prinzipiell jede *CCDB*-Kookkurrenz über das COSMAS-System

5 Vgl. „Generalisierungen über Ähnlichkeitsbeziehungen nennen wir allgemein *Präferenzrelationen*.“ bei Kupietz und Keibel (2009b b: 45); Hervorhebung im Original.

6 Stand November 2008: Dies ist ein ca. 2,2 Milliarden laufende Textwörter umfassendes, auf DeReKo basierendes virtuelles Korpus *ccdb-2007*.

online rekonstruiert und nachvollzogen werden. Die Kookkurrenzdatenbank ist im Sinne eines „gläsernen Labors" teilweise online zugänglich unter <http://corpora.ids-mannheim.de/ccdb/>.

## 4.2. Paarweiser Vergleich von Kookkurrenzprofilen

Vergleicht man synoptisch die Kookkurrenzprofile zweier Wörter, die intuitiv als „ähnlich" empfunden werden, z.B. die eines synonymischen Paares, so stellt man fest, dass sich auch diese Kookkurrenzprofile teilweise ähneln, sie weisen leicht erfassbare partielle Übereinstimmungen in ihren Kookkurrenzclustern auf. In der Abbildung 2 sieht man beispielsweise, dass die Kookkurrenzprofile von *grinsen* (links) und *lächeln* (rechts) bereits in ihren ersten Zeilen vier Bereiche mit teilweiser Überlappung haben, wobei die Gemeinsamkeiten die Kookkurrenzen mit den Wörtern *sagt*, *verschmitzt*, *freundlich* und *verlegen* betreffen.

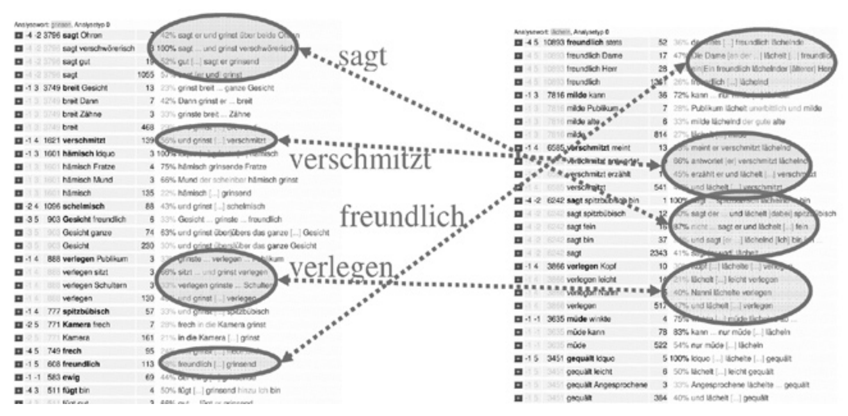



Abb. 2: Synoptischer Vergleich der Kookkurrenzprofile von *grinsen* und *lächeln*

Im Umkehrschluss wird angenommen, dass man Objekte, deren Kookkurrenzprofile messbare Überlappungen aufweisen, als *in ihrem Gebrauch ähnlich* bezeichnen kann. Es bleibt eine offene Forschungsfrage, wie ein optimales Ähnlichkeitsmaß für Objekte mit der Komplexität von Kookkurrenzprofilen konzipiert sein soll. Das derzeit in der *CCDB* verwendete Maß beruht auf dem Termfrequenz/Suchwortdichte-Ansatz (*tf-idf*, s. zum Beispiel Salton 1989: 348). Einen ersten Eindruck davon, welches Konzept von Ähnlichkeit dieses Maß operationalisiert, gewinnt man, indem man zu einem gegebenen Wort die Wörter ausgeben lässt, die mit ihren Kookkurrenzprofilen dem Wort gemäß diesem Maß am stärksten ähneln. Als ein Beispiel zeigt Abbildung 3 die Liste der 20 Wörter, deren Kookkurrenzprofile in der *CCDB* dem Kookkurrenzprofil des Wortes *Hindi* am ähnlichsten sind.

1. Chinesisch
2. Englisch
3. Spanisch
4. Türkisch
5. Urdu
6. Portugiesisch
7. Japanisch
8. Arabisch
9. Italienisch
10. Landessprache
11. Polnisch
12. Französisch
13. Muttersprache
14. Griechisch
15. Hebräisch
16. Ungarisch
17. Amtssprache
18. Tschechisch
19. Russisch
20. Niederländisch

Abb. 3: Wörter mit Kookkurrenzprofilen, die dem von Hindi am ähnlichsten sind

Während die intuitiv leicht nachvollziehbare Ähnlichkeit dieser Wörter – im Sinne von semantischer Nähe – zu *Hindi* unspektakulär ist, da es sich überwiegend um Fälle von Kohyponymie handelt und manche dieser Beziehungen möglicherweise bei Priming-Experimenten signifikant häufig elizitiert würden, ist die Situation bei dem Wort *Charakteristikum* (s. Abbildung 4) spürbar komplexer und jede Vorwegnahme von Ergebnissen experimentalpsychologischer Untersuchungen erscheint hier fragwürdig und unseriös.

1. Merkmal
2. Eigenheit
3. Eigenschaft
4. Eigenart
5. Ausprägung
6. Charakteristik
7. Anliegen
8. Element
9. Besonderheit
10. Charaktereigenschaft
11. Ausformung
12. Stilelement
13. Kriterium
14. Charakterzug
15. Stilmittel
16. Parameter
17. Charakter
18. Eigentümlichkeit
19. Ereignis
20. Spielart

Abb. 4: Wörter mit Kookkurrenzprofilen, die dem von Charakteristikum am ähnlichsten sind

Im Allgemeinen kann man den folgenden Zusammenhang erwarten: Je facettenreicher die (denotativen und konnotativen) semantischen Strukturen sind, die mit einem *Lexem* (d.h. *nicht* mit einem seiner Denotate) typischerweise assoziiert werden, desto heterogener und divergenter ist die Menge der Lexeme, die ihm *in ihrem Gebrauch* ähnlich sind. Diese Vermutung wird im Folgenden weiter erkundet.

Zusammenfassend soll an dieser Stelle betont werden, dass eine begleitende interdisziplinäre sprachpsychologische Untersuchung von Ähnlichkeitsrelationen und -strukturen zwischen *Lexemen* für eine adäquate korpuslinguistische Modellierung lexikalischer Semantik unverzichtbar ist, und dass sie eine mindestens ebenso große Herausforderung wie die meistens leider allein diskutierte mathematisch-statistische Operationalisierung darstellt. Sobald die Sprache selbst, und nicht z.B. die durch sie transportierte Information empi-

risch untersucht und erschlossen werden soll, muss der Bezug zu den kognitiv-psychologischen Wurzeln von Sprache nicht nur umrissen werden, sondern in der gesamten Forschungsmethodik durchgehend aufrechterhalten bleiben.

## 4.3. Vergleich von mehreren Kookkurrenzprofilen

Die im Abschnitt 4.2 skizzierten Überlegungen zur partiellen Ähnlichkeit zweier Lexeme (basierend auf partieller Ähnlichkeit ihrer Kookkurrenzprofile) kann man mit dem folgenden Gedankenexperiment fortführen: Wie oben angedeutet, ist es möglich, zu einem gegebenen Lexem λ eine Menge von Lexemen Σ zu nennen, die dem Lexem λ in seinem Gebrauch ähnlich sind. Dabei können diese Lexeme auf sehr unterschiedliche Art zu λ ähnlich sein, d.h. sie können jeweils einen sehr spezifischen Verwendungsaspekt mit λ teilen. Hat man einen sehr großen Vorrat an „verschiedenartig" ähnlichen Lexemen in Σ zur Verfügung, so „berühren" diese das Lexem λ von jeweils unterschiedlichen Seiten, sie bringen jeweils andere Verwendungsaspekte von λ zum Ausdruck. Im Idealfall wäre man in der Lage, zu *jedem* bekannten Verwendungsaspekt des Lexems λ mehrere Lexeme aus der Vorratsliste Σ zu nennen, die eben diesen Verwendungsaspekt mit λ teilen. Wäre es dann zusätzlich möglich, die Lexeme aus Σ so zu Teilmengen zu gruppieren, dass jede Teilmenge einem der sich abzeichnenden Verwendungsaspekte entspricht, so könnte man sagen, dass diese Teilmengen das Lexem λ Aspekt für Aspekt „abgetastet", „kartiert" und dadurch letztendlich bestimmt haben – etwa so, wie man mit geschlossenen Augen durch Abtasten einen Gegenstand zu erfassen versucht.
Etwas formaler ausgedrückt: Über die Analyse der Struktur der Kookkurrenzprofile und des multidimensionalen Vektorraumes, den diese Kookkurrenzprofile vermöge einer Ähnlichkeitsmetrik aufspannen, gelangt man zu Erkenntnissen über die Struktur bestimmender Verwendungsaspekte von lexikalischen Einheiten, insbesondere über die Struktur ihrer lexikalisch-semantischer Relationen. Im Sinne der im Abschnitt 2 zitierten Maxime „[…] sich dem Untersuchungsgegenstand Sprache mit möglichst wenigen Vorannahmen über diesen Gegenstand selbst zu nähern", wird dazu, d.h. zur Analyse bzw. Dimensionsreduktion des Vektorraumes, der Ansatz von neuronalen Netzen – in Form von selbstorganisierenden Merkmalskarten – aufgegriffen. Abbildung 5 zeigt eine mögliche lexikalische Merkmalskarte für das Lexem *Quark*.

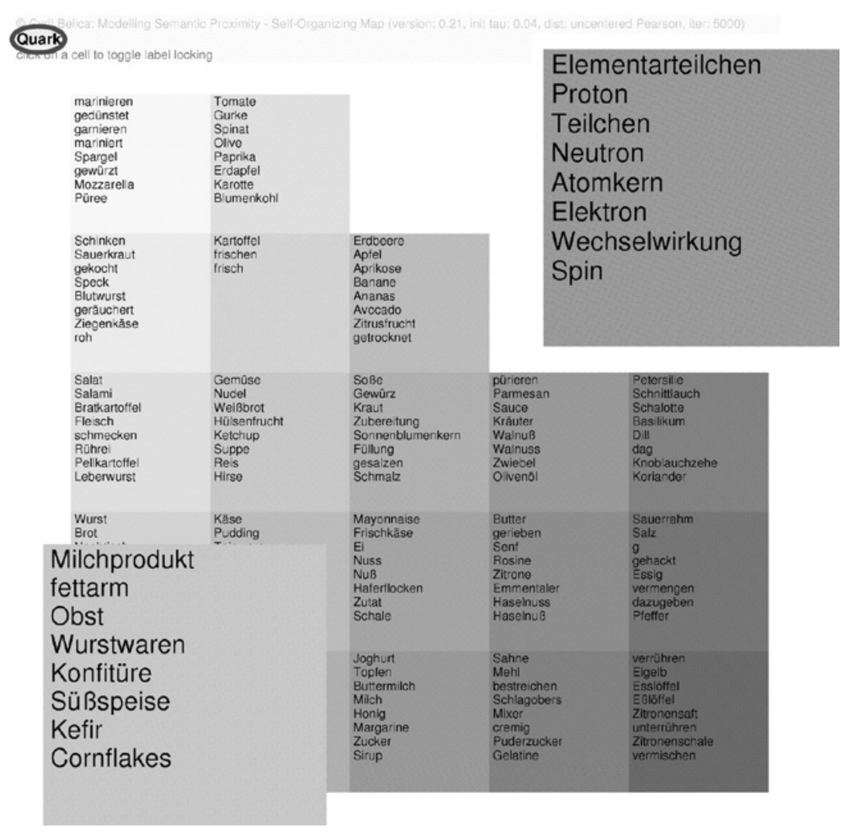



Abb. 5: Selbstorganisierende lexikalische Merkmalskarte[7] für das Lexem *Quark*

Diese Merkmalskarte enthält ausschließlich Lexeme, deren Kookkurrenzprofile dem Kookkurrenzprofil des Lexems *Quark* besonders ähnlich sind. Sie wurden mittels Selbstorganisation so angeordnet, dass die topografische Nähe der Lexeme auf der Fläche möglichst der Ähnlichkeit ihrer Kookkurrenzprofile untereinander entspricht. So sind z.B. die Kookkurrenzprofile von *Konfitüre* und *Süßspeise* nicht nur beide dem Kookkurrenzprofil von *Quark* ähnlich, sondern sie sind – so folgert man – auch untereinander ähnlich, da sie sich im Verlauf der Selbstorganisation nahe beieinander, in demselben Quadrat (s. linke untere Ecke der Merkmalskarte) stabilisiert haben. In der größten Entfernung davon, im Eckquadrat rechts oben, findet man allerdings Lexeme, die

7 Die Vergrößerung der zwei im Text diskutierten Quadrate in dieser Abbildung ist nicht ein Bestandteil der automatischen Visualisierung und wurde zwecks besserer Übersichtlichkeit nachträglich von Hand vorgenommen.

man zwar auch leicht mit dem Lexem *Quark*, kaum aber mit den Lexemen *Konfitüre* oder *Süßspeise* assoziiert, z.B. *Proton*, *Neutron*, *Elektron*. Es liegt daher nahe anzunehmen, dass das Lexem *Quark* in mindestens zwei[8] unterschiedlichen globalen, situativen Kontexten gebraucht wird, die hier verkürzt als *Milchprodukt*-Kontext und *Elementarteilchen*-Kontext bezeichnet werden[9]. In Anlehnung an die Argumentation im Abschnitt 2, wonach man „den vermöge gradueller Ähnlichkeit zwischen Situationen und Kontexten entstandenen unscharfen Konzeptualisierungen von sprachlichen Phänomenen […] mit Hilfe von A-posteriori-Klassifikationen eine Pseudo-Identität verleihen kann (um sie symbolisch referenzierbar zu machen)", würde man z.B. aus lexikografischer Sicht das in diesem trivialen Beispiel „aufgedeckte" Phänomen möglicherweise als Homonymie klassifizieren. Die Maxime, sprachsystembezogene *Generalisierungen* – z.B. die Abbildung von emergenten Phänomenen auf sprachsystemische Kategorien – möglichst erst nachgelagert, *a posteriori* durchzuführen, gehört zu den Grundpfeilern der in diesem Beitrag vorgestellten Methodik.

Im Allgemeinen ist die linguistische Interpretation von selbstorganisierenden lexikalischen Merkmalskarten eine nichttriviale Herausforderung. Vachková und Belica (2009) arbeiten die Methodik einer im Diskurs verankerten semiotischen Interpretation derartiger Merkmalskarten aus und zeigen, dass die emergenten Zeichen intersubjektiv auf das mentale Lexikon bezogen werden können („'re-anchor[ing]' [of] the representamen to a substitute semiotic object in […] mental lexicon."). Sie stellen fest, dass der zu interpretierende Signifikant kein einzelnes Lexem, sondern „a blurred group of topologically close lexemes" ist und charakterisieren dessen semiotisches Objekt als „the quality of the common pair-wise relationship between its constituent lexemes". Daraus folgern sie:

> Since the quality of the underlying relationship – in the sense of 'explication' – is available neither in the corpus nor in the self-organizing map, an attempted semiosis, i.e., rendering the supersign meaningful, requires that its emergent semiotic object be sought in the point of intersection of the individual semiotic objects resulting from their interaction within the mental lexicon due to their connotative potential.

---

8 Auf einen weiteren globalen Kontext des Lexems *Quark*, nämlich *Unsinn*, kann hier aus Platzgründen nicht näher eingegangen werden.

9 Es ist nur ein Zufall, obgleich ein sehr willkommener, dass die zwei diskutierten Quadrate jeweils auch das Lexem beinhalten, das direkt zur Kennzeichnung der zwei erwähnten globalen Kontexte verwendet werden konnte, nämlich *Milchprodukt* und *Elementarteilchen*.

Die im Anhang B abgebildete, teilweise interpretierte und annotierte lexikalische Merkmalskarte für das Wort *Glas* soll die hier skizzierte Gedankenführung illustrieren.

## 5. Verwandte Fragestellungen und Anwendungsszenarien

Im Vorangegangenen wurde eine aus mehreren Schritten bestehende Methodik dargestellt. In diesem Abschnitt werden drei Themengebiete skizziert, in denen die Ergebnisse bestimmter Teile dieser Methodik verwendet werden können.

### 5.1. Synonymie und Plesionymie

Erstellt man eine kombinierte lexikalische Merkmalskarte für ein Paar von semantisch verwandten Lexemen, indem man alle Lexeme, die mindestens einem der beiden Lexeme ähnlich sind, dem Prozess der Selbstorganisation gemeinsam unterzieht, so treten die Gemeinsamkeiten und Unterschiede in der Verwendung einzelner Wörter des untersuchten Wortpaares deutlich hervor. Diese Analysemethode (*Contrasting Near Synonyms*, Belica 2006) kann u.a. zum Aufspüren und Dechiffrieren von feinen Bedeutungsdifferenzen bei Quasisynonymen verwendet werden (vgl. Vachková / Belica, 2009 und Schmidts Analyse von *genau* in Vachková / Schmidt / Belica 2007: 18-20).
Abbildung 6 zeigt eine kombinierte lexikalische Merkmalskarte für das Wortpaar *Einsamkeit/Zweisamkeit*. Die Grauschattierung des Hintergrunds einzelner Quadrate entspricht dem Anteil, in dem die darin aufgelisteten Lexeme einem der kontrastierten Wörter im Gebrauch ähnlich sind, von weiß (entspricht dem globalen Kontext von *Einsamkeit*), über hellgrau (deutet auf globale Kontexte hin, mit denen sowohl *Einsamkeit* wie auch *Zweisamkeit* assoziiert werden könnten) bis dunkel (Situationen, die man typischerweise mit *Zweisamkeit*, nicht aber mit *Einsamkeit* verbindet).[10]

10 Die Wahl der Grauschattierung in dieser Publikation ist drucktechnisch bedingt. Im Original wird zur Visualisierung des Ähnlichkeitsgrades in den einzelnen Quadraten eine unterschiedliche Farbgebung für die Ausgangswörter (gelb bzw. rot) verwendet.

© Cyril Belica: Modelling Semantic Proximity - Contrasting Near-Synonyms (version: 0.17, init tau: 0.4, dist: x, iter: 5000)

Einsamkeit Zweisamkeit

| | | | | |
|---|---|---|---|---|
| Ehe | Familienleben | Zusammensein | Wonne | innig |
| Kleinfamilie | Umarmung | Lebensabend | Gegenwelt | Innigkeit |
| ehelich | Flitterwochen | schönen | nachhängen | Männerfreundschaft |
| außerehelich | Küssen | sorgenfrei | schön | Zwiegespräch |
| ehelichen | Verliebte | Kindheit | schrecklich | platonisch |
| Liebesbeziehung | Flirt | Herzblatt | sonderbar | Liebesszene |
| Liaison | Liebesnacht | unvergesslich | Kindheitserinnerung | Hassliebe |
| Ehejahr | schenken | unvergeßlich | wunderbar | Haßliebe |
| unzertrennlich | Liebesglück | Beschaulichkeit | Freundschaft | Heldentum |
| entfremdet | Venus | Abwechslung | Erlösung | Sexualität |
| Liebesspiel | | Nachtruhe | Privatheit | unerfüllbar |
| entfremden | | Abgeschiedenheit | immerwährend | Erotik |
| Beziehungskrise | | Geselligkeit | Gemeinsamkeit | Selbstverwirklichung |
| symbiotisch | | Idylle | Ganzheit | Unmittelbarkeit |
| | | Muße | Feindschaft | pubertär |
| | | Behaglichkeit | Seligkeit | voyeuristisch |
| Erniedrigung | Sinnlosigkeit | Weite | Alleinsein | Zärtlichkeit |
| Schmerz | Vergänglichkeit | Erhabenheit | Sehnsucht | unerfüllt |
| Qual | Tragik | Sinnbild | Intimität | Liebe |
| Entbehrung | Endlichkeit | unendlich | Zartheit | unstillbar |
| Demütigung | thematisiert | Unendlichkeit | Schwermut | Leidenschaft |
| Leid | Unmöglichkeit | Gefühlswelt | Empfindsamkeit | Verliebtheit |
| Mühsal | Fremdsein | Metapher | Geborgenheit | Verlangen |
| Pein | Dasein | Freiheit | Vergeblichkeit | Zuneigung |
| Not | Ausweglosigkeit | Verlassenheit | Zerrissenheit | Lebenslust |
| Depression | Heimatlosigkeit | Langeweile | Melancholie | grenzenlos |
| Hunger | ausweglos | Trostlosigkeit | Trauer | Suff |
| Neurose | existentiell | Leere | Enttäuschung | Gier |
| Alkoholismus | Elend | Tristesse | Auflehnung | Todessehnsucht |
| Anfall | existenziell | Stille | Gefühl | Eifersucht |
| Krankheit | Entwurzelung | Düsternis | Empfindung | Geldgier |
| Drogensucht | Selbstzerstörung | versinken | Verzweiflung | abgründig |
| Vereinsamung | Orientierungslosigkeit | Hoffnungslosigkeit | Resignation | Wut |
| Isolation | Erstarrung | Ohnmacht | Verbitterung | Selbstmitleid |
| Teufelskreis | Vereinzelung | Hilflosigkeit | Aggression | Bitterkeit |
| Armut | Perspektivlosigkeit | Sprachlosigkeit | Frust | Scham |
| Verwahrlosung | Erschöpfung | Apathie | breitmachen | Selbsthaß |
| Verarmung | Überforderung | Frustration | Fassungslosigkeit | Selbsthass |
| Befindlichkeit | Aussichtslosigkeit | Ratlosigkeit | gewichen | abgrundtief |
| Verelendung | Desorientierung | Niedergeschlagenheit | umschlagen | Angst |

Abb. 6: Lexikalische Merkmalskarte für das Wortpaar *Einsamkeit/Zweisamkeit.*

## 5.2. Lexikografie und Lexikologie

Ein erstes Anwendungsgebiet, das sich aufgrund des bereits Gesagten anbietet, ist die korpuslinguistisch abgesicherte Lexikografie. Der oben erwähnte emergente Charakter der lexikalisch-semantischen und lexikalisch-syntaktischen Strukturen (vgl. z.B. Bybee 1998: 421 in Anlehnung an Hopper 1987) sowie die fruchtbare Verquickung von Welt- und Sprachwissen (vgl. Adamska-Sałaciak 2006: 47) sind zwei Aspekte, die die moderne Lexikografie besonders betont. Sie sind zugleich als zwei bedeutende Analyseperspektiven bei der Bedeutungserfassung zu nennen: Die oben vorgestellte korpusanalytische Methodik lenkt den Lexikografen in die richtigen Bahnen immer dann,

wenn er in Gefahr läuft, Wörter als feste Entitäten von Form und Bedeutung wahrzunehmen, bzw. diese auf abstrakte Art und Weise unabhängig vom Kontext zu umschreiben (anstatt sie z.B. mit semiotisch aussagekräftigem Wortmaterial zu belegen).

Der kontext- bzw. diskursbedingte Facettenreichtum spielt auch bei der Auswahl von Synonymen oder anderen zum Stichwort äquivalenten Ausdrücken eine eminente Rolle, von der die Qualität der lexikografischen Erfassung eines Wortes (bzw. einer Wortkombination) abhängt: Der Kontrast zwischen der lexikografischen Bearbeitung, die sich nur unmittelbar und hermeneutisch auf Korpusbelege stützt, und dem hier angedeuteten Ansatz, der sich die hier vorgestellte Auffassung der lexikalisch-semantischen Beziehungen zu eigen macht, kann als eine Triebkraft der lexikografischen Arbeit verstanden werden. Als Beispiel kann eine neu zu entwerfende konzeptuelle Auffassung von antonymischen/synonymischen Beziehungen dienen, bei der die skalare Beschaffenheit der Gegensätze diskursgebunden betrachtet wird (vgl. Vachková, 2010). Dies ist eine Herausforderung nicht nur für die Lexikografie, sondern auch für die korpusbasierte Lexikologie.

Hiermit ist bei der Anwendung der korpuslinguistischen Methodik die Frage nach der Vernetzung eines Wortes im Rahmen der paradigmatischen Beziehungen und nach der Eruierung der Wortbedeutung auf einer qualitativ neuen Ebene zu beantworten. Daraus resultiert eine Auffassung von lexikalischen Relationen, die nicht von der a priori verankerten Kategorisierung ausgeht: Das Einbeziehen von Kookkurrenzprofilen in die lexikografische Arbeit bedeutet, einen schnellen und übersichtlichen Zugriff auf eine assoziativ signifikante Diskursmenge als Orientierungsskelett zur Verfügung zu haben.

Die laufende Forschung, die sich mit der linguistischen Interpretation von lexikalischen Merkmalskarten befasst, deutet darauf hin, dass der Prozess, in dem sich das kommunikative Potential der Sprache realisiert, als emergente Struktur visualisierbar ist. Die am Prozess der Selbstorganisation beteiligten Kookkurrenzprofile zeigen sich dabei sowohl bedeutungskonstituierend wie auch bedeutungsrelativierend. Vachková (in Vorbereitung) merkt an:

Die Erfahrung, dass eine diskursbasierte Interpretation von Assoziationsnetzen z.B. die lexikografische Bearbeitung von schwierigen Abstrakta besser steuert als eine abstrakte Bedeutungsumschreibung, erhärtet die Schlussfolgerung von Schwanenflugel: „When abstract and concrete words are placed in highly supportive contexts [sprich: *in Clustern von Kookkurrenzprofilen*], the preactivation of contextual knowledge [sprich: durch *Assoziationen evoziertes Diskurswissen, das an Weltwissen gekoppelt ist*] should override processing [*bzw. priming*] difficulties for abstract words".

Da die der Methodik zugrundeliegende iterative Kookkurrenzanalyse neben binären Wortrelationen auch ganze phrasale Muster erfasst, eröffnen sich außerdem der Lexikografie – auf sehr große Korpora gestützt – nahezu unerschöpfliche Möglichkeiten zur Inventarisierung, Klassifikation und Beschreibung usueller Wortverbindungen – z.B. Phraseologismen, Redewendungen, Sprichwörter, kommunikative Formeln, Funktionsverbgefüge – und der in ihnen vorkommenden (insbesondere paradigmatischen) Variation.

## 5.3. Didaktik

Ein drittes Themengebiet betrifft nur einen Teil der oben dargestellten Methodik: die syntagmatischen Muster. Aus der in diesem Aufsatz eingenommenen Perspektive auf den sprachtheoretischen Status usueller Wortverbindungen, wie sie sich in den syntagmatischen Mustern (s. Abschnitt 3.1) manifestieren, dass sie nämlich *psychologisch real* sind, liegt es nahe, ihren Stellenwert auch im didaktischen Kontext zu überprüfen. Unter dem Arbeitstitel *simulated late partial immersion* wird am IDS eine weitere Anwendungsmöglichkeit diskutiert mit der Zielvorstellung, ein Kompendium usueller Wortverbindungen für Deutschlernende zu konzipieren. Der Entwurf sieht vor, dass zuerst aus der Kookkurrenzdatenbank *CCDB* solche syntagmatischen Muster extrahiert werden, in denen lediglich Kookkurrenzpartner aus einem gewählten Lernervokabular vorkommen. Diese werden dann redaktionell geprüft, ggf. korrigiert und didaktisch aufbereitet, und wahlweise – auf halbautomatischem Wege – um Beispiele ihrer Verwendung im Korpus ergänzt. Das vielschichtige Zusammenspiel von Faktoren wie Sprachanlass/-situation und Lernniveau könnte je nach Zielpublikum angemessen berücksichtigt werden, da es möglich wäre, das dem Kompendium zugrundeliegende Vokabular in seinem Umfang und in seiner Zusammensetzung praktisch ohne zusätzlichen Aufwand beliebig zu variieren. Aufgrund der großen systematischen (v.a. paradigmatischen) und idiosynkratischen Variation der äußeren Formen dieser Kookkurrenzverbindungen ist es allerdings nicht einfach, für sie eine Art kanonische Grundform einzuführen und ggf. zu operationalisieren, die ihr jeweils eine für didaktische Zwecke hinlänglich fassbare Identität verleihen würde.
Heringer (2007) berichtet von einem gemeinsamen Experiment in diesem Sinne unter Verwendung eines erweiterten Zertifikatsvokabulars mit ca. 2500 hochfrequenten Grundformen. Er nennt diese Wortverbindungen „Valenz-Chunks" und gibt interessante Beispiele für ihre didaktisch motivierte Aufbereitung. Anhang C listet einige Beispiele für syntagmatische Muster, die in diesem Experiment verwendet wurden, in ihrer ursprünglichen, nicht aufbereiteten Form auf.

## 6. Zusammenfassung

In diesem Beitrag wurde der Begriff der *semantischen Nähe* aus einer emergentistischen Perspektive vor dem Hintergrund eines korpusgeleiteten empirisch-linguistischen Ansatzes diskutiert. Es wurde eine konsistente und für die *explorative Erforschung von Sprache mit explanatorischem Anspruch* adäquate Methodik vorgestellt, die sich ihrem Untersuchungsgegenstand mit möglichst wenigen Vorannahmen über diesen Gegenstand selbst nähert. Die Methodik wurde in vier Schritten präsentiert:

Auf einer *Erweiterung* des Begriffs der *Kookkurrenz* aufbauend wurden zuerst *Kookkurrenzprofile* als kondensierte Aufzeichnungen einer *psychologisch realen* syntagmatischen und paradigmatischen Einbettung lexikalischer Einheiten im Sprachgebrauch aus präferenzrelationaler Sicht konzeptuell eingeführt. Geleitet von einem weit gefassten kognitiven Prinzip der *Ähnlichkeit* – als einem graduellen Gegenkonzept zur Identität – wurde ein Vektorraum von solchen Kookkurrenzprofilen auf der Basis eines sehr großen Korpus konstruiert. Im zweiten Schritt wurde dargelegt, dass dieser Vektorraum mit Hilfe von geeigneten analytischen Methoden – den *selbstorganisierenden Merkmalskarten* – auf seine Struktur untersucht werden kann. Der dritte Schritt bestand darin zu zeigen, dass die aufgedeckten Strukturen als *emergente Zeichen* mittels einer im Diskurs verankerten *semiotischen Interpretation* zu den globalen Kontexten, in denen sich die Entstehung und Festigung denotativer und konnotativer Aspekte lexikalisch-semantischer Relationen im konkreten Sprachgebrauch ereignet, intersubjektiv zurückverfolgt werden können. Die Forderung, sprachsystembezogene *Generalisierungen* – z.B. die Abbildung von emergenten Phänomenen auf sprachsystemische Kategorien – erst nachgelagert, *a posteriori* durchzuführen, war schließlich Gegenstand des vierten Schritts.

Weitere Fortschritte und eine auf experimentelle Daten gestützte Argumentation werden von den geplanten *interdisziplinären Untersuchungen* im Schnittpunkt von Korpuslinguistik, Psycholinguistik und kognitiver Psychologie erwartet.

## Anhang A

Der folgende Ausschnitt aus den Ergebnissen der Kookkurrenzanalyse des Wortes *why* illustriert die Sprachunabhängigkeit der verwendeten Analysemethode. Diese Analyse wurde in einem ad-hoc zusammengestellten Geschäftsenglisch-Korpus mit etwa zwei Millionen laufenden Wörtern durchgeführt.

Analysewort why-en, Analysetyp 0

<table>
<tr><td>-1</td><td>-1</td><td>1805</td><td>reason one main</td><td>1</td><td>100%</td><td>reason ... main one why</td></tr>
<tr><td>-1</td><td>-1</td><td>1805</td><td>reason one</td><td>64</td><td>96%</td><td>is one reason [...] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>1805</td><td>reason main</td><td>21</td><td>90%</td><td>The|the main reason why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>1805</td><td>reason One</td><td>24</td><td>100%</td><td>One reason [...] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>1805</td><td>reason</td><td>181</td><td>100%</td><td>is one reason [...] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>1282</td><td>explain helps</td><td>23</td><td>100%</td><td>This helps [to] explain [...] why ... the</td></tr>
<tr><td>-1</td><td>-1</td><td>1282</td><td>explain may help</td><td>3</td><td>100%</td><td>may help [to] explain why</td></tr>
<tr><td>-1</td><td>-1</td><td>1282</td><td>explain may</td><td>28</td><td>96%</td><td>This may [...] explain why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>1282</td><td>explain help</td><td>13</td><td>100%</td><td>may|might help [to] explain why</td></tr>
<tr><td>-1</td><td>-1</td><td>1282</td><td>explain</td><td>113</td><td>100%</td><td>helps may|to explain [...] why</td></tr>
<tr><td>-1</td><td>-1</td><td>575</td><td>is That</td><td>78</td><td>83%</td><td>That [...] is [...] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>575</td><td>is easy It</td><td>13</td><td>92%</td><td>It is easy to see why</td></tr>
<tr><td>-1</td><td>-1</td><td>575</td><td>is easy</td><td>19</td><td>89%</td><td>It|it is [...] easy to see why</td></tr>
<tr><td>-1</td><td>-1</td><td>575</td><td>is It</td><td>21</td><td>90%</td><td>It is easy to see why the</td></tr>
<tr><td>-1</td><td>-1</td><td>575</td><td>is</td><td>393</td><td>71%</td><td>That is [... reason] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>543</td><td>explains This partly</td><td>3</td><td>100%</td><td>This [...] partly explains why</td></tr>
<tr><td>-1</td><td>-1</td><td>543</td><td>explains This</td><td>12</td><td>100%</td><td>This [partly] explains why</td></tr>
<tr><td>-1</td><td>-1</td><td>543</td><td>explains partly</td><td>7</td><td>100%</td><td>This partly explains why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>543</td><td>explains</td><td>49</td><td>100%</td><td>This explains [...] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons There are several</td><td>7</td><td>85%</td><td>There are several reasons why</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons There are</td><td>23</td><td>78%</td><td>There are several|two reasons why</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons There</td><td>24</td><td>100%</td><td>There are several|many reasons why</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons are several</td><td>9</td><td>88%</td><td>There are several reasons why</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons are</td><td>34</td><td>82%</td><td>There|there are [several|two] reasons why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons several</td><td>11</td><td>100%</td><td>There|there are several reasons why</td></tr>
<tr><td>-1</td><td>-1</td><td>528</td><td>reasons</td><td>57</td><td>100%</td><td>are ... reasons [...] why the ...</td></tr>
<tr><td>-4</td><td>-2</td><td>527</td><td>That Fund</td><td>3</td><td>100%</td><td>That is why [the] Fund</td></tr>
<tr><td>-4</td><td>-2</td><td>527</td><td>That</td><td>106</td><td>100%</td><td>That is ... why the ...</td></tr>
<tr><td>-2</td><td>-2</td><td>302</td><td>one</td><td>76</td><td>93%</td><td>is one [reason] why the ...</td></tr>
<tr><td>-1</td><td>-1</td><td>247</td><td>see hard It</td><td>1</td><td>100%</td><td>It ... hard ... see why</td></tr>
<tr><td>-1</td><td>-1</td><td>247</td><td>see</td><td>43</td><td>95%</td><td>easy|hard to see [...] why</td></tr>
<tr><td>-4</td><td>-2</td><td>235</td><td>This</td><td>75</td><td>100%</td><td>This is ... explain why</td></tr>
<tr><td>1</td><td>3</td><td>226</td><td>should no</td><td>14</td><td>92%</td><td>there is no reason why [...] should not ...</td></tr>
<tr><td>1</td><td>3</td><td>226</td><td>should be</td><td>16</td><td>93%</td><td>why [...] should [...] be</td></tr>
<tr><td>1</td><td>3</td><td>226</td><td>should</td><td>65</td><td>98%</td><td>reason why [...] should be ...</td></tr>
<tr><td>-3</td><td>-2</td><td>178</td><td>helps</td><td>24</td><td>95%</td><td>This helps to explain why ... the</td></tr>
<tr><td>1</td><td>1</td><td>162</td><td>the government</td><td>12</td><td>83%</td><td>That reason|is why [...] the [...] government ... to</td></tr>
<tr><td>1</td><td>1</td><td>162</td><td>the</td><td>368</td><td>68%</td><td>is why [...] the</td></tr>
<tr><td>-5</td><td>-4</td><td>153</td><td>There</td><td>32</td><td>100%</td><td>There is|are several reason|reasons why the ...</td></tr>
<tr><td>-2</td><td>-2</td><td>152</td><td>main</td><td>25</td><td>96%</td><td>The|the main reason why the ...</td></tr>
<tr><td>-5</td><td>-2</td><td>150</td><td>Which</td><td>18</td><td>100%</td><td>Which is|may explain why the ...</td></tr>
<tr><td>-3</td><td>-3</td><td>140</td><td>easy</td><td>22</td><td>95%</td><td>It|it is easy [to see] why</td></tr>
<tr><td>-2</td><td>-2</td><td>134</td><td>One</td><td>31</td><td>100%</td><td>One reason why the ...</td></tr>
<tr><td>1</td><td>1</td><td>129</td><td>they do</td><td>6</td><td>33%</td><td>they do not ... why</td></tr>
<tr><td>1</td><td>1</td><td>129</td><td>they</td><td>68</td><td>76%</td><td>why [...] they</td></tr>
<tr><td>-2</td><td>-2</td><td>118</td><td>to</td><td>241</td><td>53%</td><td>to [see|explain] why</td></tr>
<tr><td>-1</td><td>-1</td><td>101</td><td>understand it To</td><td>2</td><td>50%</td><td>To understand why it</td></tr>
<tr><td>-1</td><td>-1</td><td>101</td><td>understand</td><td>12</td><td>100%</td><td>To|to understand why</td></tr>
</table>

Abb. 7: Ergebnisse der Kookkurrenzanalyse von *why* (Ausschnitt)

## Anhang B




Abb. 8: Teilweise interpretierte und annotierte lexikalische Merkmalskarte zu *Glas*

# Anhang C

Dieser Anhang zeigt einige Beispiele für syntagmatische Muster aus der Kookkurrenzdatenbank *CCDB*, in denen als Kookkurrenzpartner ausschließlich Wortformen aus dem erweiterten Zertifikatsvokabular mit ca. 2500 hochfrequenten Grundformen vorkommen, vgl. Abschnitt 5.3.

war [... ein] voller [...] Erfolg
nichts zu tun [...] haben
in letzter Sekunde
kann [es ...] passieren daß
nicht [mehr] bezahlen
Licht ins Dunkel
tiefer in die Tasche [...] greifen
mit leeren [...] Händen da ...
kann nicht [...] schaden
Ich habe [... den] Eindruck dass|daß ...
die Nase [...] vorn [zu] haben
in der kalten [...] Jahreszeit
fehlt [nur] noch
steht [... ganz im] Zeichen des|der ...
ganz besonders
Das gilt [...] auch für die
allerdings [...] noch nicht ...
an allen Ecken und Enden
wird|es noch [... Jahre] dauern bis
lässt ... zu wünschen [...] übrig
schon einmal
gerade mal
paßt [...] gut … auf
eine ... Rolle [...] gespielt
funktioniert [...] nicht [...] mehr
eine Lösung [zu] finden
ganz schön
mit dem Rücken zur Wand
wer weiß
wieder auf freiem [...] Fuß angezeigt
Rede und Antwort [...] stehen
an|An erster [...] Stelle
viel Glück
Glück gehabt
Glück [im] Unglück hatte ...
ebenso [...] wie die ...
und manchmal [...] sogar
ist ... der Meinung
kommt nicht von ungefähr
an den Start [...] gehen
fast überall
in deutscher Sprache
ist [... für … nicht] geeignet
ist [...] typisch für ...

überall auf|in der Welt
das Maß der|aller Dinge
ist [noch] offen
richtet sich in erster Linie gegen|an die
für frischen [...] Wind in die
auf die Art und Weise wie ...
ein Zeichen [... zu] setzen
ein Lied [davon zu] singen
wohl [...] nicht
in die Tasche [...] greifen
in der Nacht [zum|auf] Sonntag
war ein voller [...] Erfolg
sieht [das ...] anders aus
schon mal
dient in erster Linie der ...
die|der leeren [...] Kassen
an die frische [...] Luft
in Verbindung [...] gebracht werden
erst einmal
am runden Tisch
zum Opfer [...] gefallen
leer ausgehen
will [... nichts] wissen
nicht [so] recht
in Frage [...] gestellt
ist noch [...] offen
in den ewigen [...] Frieden heimgeholt
Sprung ins kalte [...] Wasser
Es hat [... viel] Spaß [...] gemacht
unter Dach und Fach [zu] bringen
in Verbindung [zu] setzen
unter Dach [und] Fach [...] gebracht werden
auf Eis [...] gelegt
Rede und Antwort stehen
auf frischer Tat ertappt ...
in|ins Krankenhaus [eingeliefert|gebracht] werden
Es ist [... kein] Zufall daß
Professor für … [an der] Universität
in der Nacht [zum|auf] Freitag
braucht [...] nicht zu
Rechnung [zu] tragen
stieg die Zahl der ...
wird nicht [...] müde zu

Grund zum Feiern
ja oder nein
die Tür [und Tor] geöffnet
eine ganze Menge
in der Nacht [zum|auf] Dienstag
in erster Linie
eine ganz normale
macht [...] Spaß
grünes [...] Licht für ...
ernst [zu] nehmen
nicht ernst [...] genommen werden
Temperaturen von|um|bis ... Grad
eine ... Rolle [...] spielen
ins Leben [...] gerufen
unter Dach und Fach
spielt [... keine|eine ...] Rolle
zur Kasse [...] gebeten werden
Auf|auf den ersten [...] Blick
über weite [...] Strecken
zu wünschen [...] übrig
die Nase [...] vorn
nicht [...] anders als
nicht [...] nur [... sondern] auch
grünes [...] Licht [für ...] gegeben
kann [sich|mir nicht] vorstellen daß ...
noch [...] lange nicht ...
Ich glaube [...] nicht daß|dass ...
einen Beitrag [zur … zu] leisten
Kilometer [von ...] entfernt
mehr oder weniger
nur noch
Erfahrungen [... zu] sammeln
nicht wirklich …
in den eigenen [...] vier [...] Wänden
steht ganz im Zeichen des|der ...
schon [...] lange nicht mehr
wohl [...] auch
sich|mir nicht [...] vorstellen daß
erst einmal
Antworten [auf die ...] Fragen
miteinander [...] verbunden sind
nicht gefallen [...] lassen
legt [...] Wert darauf|auf die
grünes Licht [für ...] geben

## Literaturverzeichnis

Adamska-Sałaciak, Arleta (2006): *Meaning and the Bilingual Dictionary*. Polish Studies in English Language and Literature. Hg. Jacek Fisiak. Peter Lang.

Belica, Cyril (1995): *Statistische Kollokationsanalyse und -clustering.* Korpuslinguistische Analysemethode. Institut für Deutsche Sprache, Mannheim.

Bybee, Joan (1998): *The emergent lexicon*. CLS 34: The panels. Chicago Linguistics Society. 421-435.

Deerwester, Scott / Dumais, Susan T. / Furnas, George W. / Landauer, Thomas K. / Harshman, Richard (1990): *Indexing by latent semantic analysis.* In: Journal of the American Society for Information Science, 41(6), 391-407.

Heringer, Hans Jürgen (2007): *Deutsch lernen mit Valenz-Chunks*. In: Zeitschrift für Angewandte Linguistik 47, 2007, 3-16.

Hopper, Paul (1987): *Emergent grammar*. Berkeley Linguistics Conference (BLS), 13:139-157.

Keibel, Holger / Kupietz, Marc (2009): *Approaching grammar: Towards an empirical linguistic research programme.* In: Minegishi, Makoto / Kawaguchi, Yuji (Eds.): Working Papers in Corpus-based Linguistics and Language Education, No. 3 (pp. 61-76). Tokyo: Tokyo University of Foreign Studies (TUFS).

Keibel, Holger / Belica, Cyril (2007): *CCDB: A Corpus-Linguistic Research and Development Workbench.* In: Proceedings of 4th Corpus Linguistics 2007, Birmingham. <http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf>

Keibel, Holger / Kupietz, Marc / Belica, Cyril (2008): *Approaching grammar: Inferring operational constituents of language use from large corpora.* In: Šticha, František / Fried, Mirjam (eds.): Selected contributions from the conference Grammar and Corpora, Sept. 25-27, 2007, Liblice, Czech Republic. Prague: ACADEMIA, S. 235-242.

Kupietz, Marc (2008): *DeReKo durchbricht Drei-Milliarden-Grenze*. Sprachreport 2/2008: p. 28 - Mannheim: Institut für Deutsche Sprache

Kupietz, Marc / Keibel, Holger (2009a): *The Mannheim German Reference Corpus (DeReKo) as a basis for empirical linguistic research.* In: Minegishi, Makoto / Kawaguchi, Yuji (Eds.): Working Papers in Corpus-based Linguistics and Language Education, No. 3 (pp. 53-59). Tokyo: Tokyo University of Foreign Studies (TUFS).

Kupietz, Marc / Keibel, Holger (2009b): *Gebrauchsbasierte Grammatik: Statistische Regelhaftigkeit.* In: Konopka, Marek/Strecker, Bruno (Hrsg.): Deutsche Grammatik - Regeln, Normen, Sprachgebrauch. S. 33-50 - Berlin/New York: de Gruyter, 2009. (Jahrbuch des Instituts für deutsche Sprache 2008)

Kohonen, Teuvo (1990): *The Self-Organizing Map.* In: New Concepts in Computer Science: Proc. Symp. in Honour of Jean-Claude Simon, p. 181-190. Paris, 1990. AFCET.

Perkuhn, Rainer (2007): *Systematic Exploration of Collocation Profiles*. In: Proceedings of $4^{th}$ Corpus Linguistics 2007, Birmingham. <http://corpus.bham.ac.uk/corplingproceedings07/paper/132_Paper.pdf

Schwanenflugel, Paula J. / Gaviska, David C. (2005): *Psycholinguistic Aspects of Word Meaning.* In: Cruse, Alan D. / Hundsnurscher, Franz / Job, Michael/ Lutzeier, Peter (Hrsg.): *Lexikologie. Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. 2. Halbband. Handbücher für Sprache und Kommunikation. HSK 21.2 S. 1735-1748. -Berlin, New York: Walter de Gruyter.

Salton, Gerard (1989): *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley, 1989.

Vachková, Marie (2010): *Zur Erforschung und Erfassung der diskursgebundenen semantischen Kontraste auf der Grundlage des SOM–Modells.* In: Tématické číslo Germanistica Pragensia XX. AUC Philologica 2. Praha. Karolinum. 193-208. ISBN 978-80-246-1799-2

Vachková, Marie (2007): *Adjektive auf -bar in kontrastiver und korpuslinguistischer Sicht. Eine metalexikographische Betrachtung.* Linguistica Pragensia 2, S. 57-74. Praha.

Vachková, Marie / Belica, Cyril (2009): *Self-Organizing Lexical Feature Maps. Semiotic Interpretation and Possible Application in Lexicography*. In: IJGLSA 13, 2 [Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis, Rauch, Irmengard and Seymour, Richard K., (eds.). – Berkeley: IJGLSA/University of California Press], pp. 223-260. ISSN 1087-555.

Vachková, Marie / Marková, Věra / Belica, Cyril (2008): *Korpusbasierte Wortschatzarbeit im Rahmen des fortgeschrittenen Germanistikunterrichts*. Zielsprache Deutsch, 3/2008.

Vachková, Marie / Schmidt, Marek / Belica, Cyril (2007): *Prager Wanderungen durch die Mannheimer Quadrate*. In: Sprachreport Sonderheft/März 2007. Auslandskooperationen des Instituts für Deutsche Sprache. S. 16-21 - Mannheim: 2007.

## Internetadressen[11]

Belica, Cyril (2007): *Kookkurrenzdatenbank CCDB - V3. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. <http://corpora.ids-mannheim.de/ccdb/>

Belica, Cyril (2006): *Modellierung semantischer Nähe: Kontrastierung von nahen Synonymen.* Korpusanalytische Methode. <http://corpora.ids-mannheim.de/ccdb/>

11 kontrolliert am 15. November 2008

Belica, Cyril (2005): *Modellierung semantischer Nähe: Analyse und topografische Visualisierung von Verwendungsaspekten in Self-Organizing-Maps.* Korpusanalytische Methode. <http://corpora.ids-mannheim.de/ccdb/>

DeReKo (2008): *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Sprache 2008-II*, Institut für Deutsche Sprache, Release vom 18.08.2008. <http://www.ids-mannheim.de/kl/projekte/korpora/archiv.html>

Institut für Deutsche Sprache (1991-2008): *COSMAS I/II (Corpus Search, Management and Analysis System).* <http://www.ids-mannheim.de/cosmas2/>

# Zur Überprüfung von Schreibleistungen an Schulübergangen am Beispiel Mittelschule-Oberschule in Südtirol[1]

**Annemarie Saxalber – Freie Universität Bozen**
**Rita Gelmi – Freie Universität Bozen**

## Abstract

Im Beitrag werden Prüfungskultur und Schreibleistungen an der Nahtstelle Mittel- / Oberschule der deutschen Schule in Südtirol in den Blick genommen[2]. Wir konzentrieren uns dabei auf die schriftliche Abschlussprüfung der Mittelschule (8. Schulstufe) im Fach Muttersprache Deutsch und im Fach Zweitsprache Italienisch. Es geht vor allem um eine kritische Beleuchtung der in den vorliegenden Dokumenten sichtbaren Prüfungskultur und deren innerer Stimmigkeit. Wir fragen danach, wie klar, transparent und kohärent in der Aufgabenstellung und in der Bewertung bei Prüfungen umgegangen wird und inwieweit dabei auch Sprachlernbezüge, die für den Start in die nächste Bildungsstufe aufschlussreich sind, hergestellt werden. Am Beispiel der schriftlichen Prüfungsaufgaben aus Deutsch/Muttersprache und Italienisch/Zweitsprache soll die Bildungsdiskussion um einen koordinierten Sprachenunterricht ein Stück weitergebracht werden.

## 1. Einleitung

Die Prüfungshandlungen bei Abschluss einer Schulstufe greifen in das Selbstverständnis eines jungen Menschen meist mehr ein, als es den Erwachsenen bewusst ist: Klaffen die gewonnene Selbsteinschätzung und die in einem Prüfungsakt festgehaltene Fremdbeurteilung zu Befähigungen arg auseinander, so wird die daraus resultierende persönliche Verunsicherung oft in die nächste Ausbildungsstufe mitgenommen. Dies kann den Start unvorteilhaft beeinflussen, zumal gerade in der Übertrittsphase auch die individuelle Betreuung der Schülerin/des Schülers vom Wechsel betroffen ist.

1 Der vorliegende Beitrag beruht auf einem Input der Autorinnen anlässlich der Herbsttagung für die Schulführungskräfte an der Deutschen Schule in Südtirol, Sekundarstufe I und II, organisiert vom Deutschen Schulamt in Brixen/Cusanus Akademie am 22.10.08. In der Zwischenzeit haben sich die gesetzlichen Unterlagen zu den Abschlussprüfungen verändert. Dies wurde im Artikel berücksichtigt. Es handelt sich hier nicht um eine repräsentative Untersuchung; die Beispiele sollen die Beobachtungen, die die Autorinnen in vielfältigen schulischen Prüfungssituationen machen konnten, veranschaulichen. Die Autorinnen danken den Schulen, die Unterlagen zur Verfügung stellten.

2 Auf die Abschlussprüfung/Mittelschule (8. Schulstufe) aus Englisch wird aus organisatorischen Gründen nur fallweise hingewiesen.

Der Eintritt in einen neuen Schultyp ist für einen jungen Menschen, für seine soziale, motivationale, kognitive wie für die Sprachentwicklung immer eine Herausforderung: Neue organisatorische Rahmenbedingungen, Bezugspersonen und Kommunikationskulturen drängen sich auf und nötigen der Schülerin/dem Schüler einiges an Anpassung und Neupositionierung ab. Damit der Übergang von der alten in die neue Schule gut gelingt, ist es wichtig, dass die Leistungsmessung am Ende der Abgeberschule – wie wohl sie auch eine punktuelle Standortbestimmung ist – in einem fortdauernden individuellen Bildungsprozess eingebettet gesehen wird und bleibt. Die Gestaltung der Überprüfung ist so zu organisieren, dass Transparenz gewährleistet ist. Transparenz unterstützt die Vergleichbarkeit und das Begründen bei Selektionsentscheidungen – wichtig für die Akzeptanz der Lehrerentscheidungen wie der Institution – und kann dazu beitragen, interdisziplinäre Vernetzungen im Fachlichen wie Pädagogischen sichtbarer zu machen. Darüber hinaus hilft Transparenz der Schülerin/dem Schüler ihren/seinen Bildungserfolg mit Distanz zu betrachten, mit Kriterien und Bewertungen souverän umzugehen, die Beurteilung innerlich anzunehmen und darauf aufbauend neue selbsttätige Schritte für die Zukunft zu setzen. Die Berücksichtigung von Transparenz zieht eine innere Koordinierung und Abstimmung der Bewertungsformen und -inhalte und klare Anleitungen mit sich. Zwischen Prüfungsgegenstand, Anforderungen, Bewertungskriterien, Aufgabenformaten, d.h. Prüfungsaufgaben oder Themenstellungen, besteht ein Zusammenhang, dem wir im Folgenden exemplarisch nachgehen wollen. Es geht vor allem um eine kritische Beleuchtung der in vorliegenden Dokumenten sichtbaren Prüfungskultur und deren innerer Stimmigkeit. Wir fragen danach, wie klar, transparent und kohärent in der Aufgabenstellung und in der Bewertung bei Prüfungen umgegangen wird und inwieweit darin auch sprachenübergreifende und lernerorientierte Bezüge Berücksichtigung finden. Bei allem ist zu bedenken, dass die gesetzlichen Vorgaben in Italien bzw. Südtirol seit zwei Jahren einer starken Veränderung unterzogen wurden. Die Fort- und die Ausbildung sind sehr gefordert, den neuen Anforderungen gerecht zu werden.
In einem ersten Abschnitt (siehe Kapitel 2) wird mit Überlegungen zur Sprachentwicklung und zum Sprachenlernen im Mittelschulalter der pädagogische und sprachdidaktische Rahmen abgesteckt, dem folgt eine Hinwendung zur Ausformung der Kompetenzenorientierung an Südtirols Schule und die Sichtbarmachung derselben in den gesetzlichen Vorgaben für die Abschlussprüfung. In Kapitel 3 schränken wir das Thema: „Überprüfen von Schreibleistungen“ auf wenige Aspekte ein: Wir hinterfragen die impliziten Anforderungsprofile an die Kandidaten/innen, die hinter (beispielhaft ausgewählten) Themenstellungen stehen, indem wir auch auf die Lehrerkommentare (verbale Begründungen) zu den Schülertexten eingehen. In den

Schlussüberlegungen (Kapitel 4) wird nochmals unterstrichen, dass Prüfungskultur und Schnittstellendidaktik zusammenhängen und dass qualitative Veränderungen in der Herkunftsschule nachhaltig auf die nächste Schulstufe einwirken.

## 2. Sprachentwicklung in der Zeit der Mittelschule

In der Zeit des Übergangs von der Grund- zur Mittelschule, also im Alter von 10 -11 Jahren lässt sich beim Kind eine Weiterentwicklung des Sprachverhaltens, die mit der kognitiven, psychischen und sozialen Reifung einher geht, beobachten. Da ist z.B. die starke Orientierung an Bezugspersonen und deren Werten zu nennen, die unter kommunikativem Aspekt gesehen sich auch in Imitation (z.B. Gestik, Wortwahl, Intonation) abbilden kann. Neugierde und eine große Wissensbehaltfähigkeit korrelieren mit der Lust an Außen- und Weltzuwendung, die den Heranwachsenden voll erfasst. Der (schulische) Wortschatz wächst stark an, die grammatischen Strukturen werden ausgebaut, die Merkmale einer formalisierteren Sprache, welche den Weg zu einer entfalteten Schriftlichkeit ebnet, eingeübt. Im Allgemeinen in der zweiten Hälfte der Mittelschulzeit besonders sichtbar nimmt die Fähigkeit zur Ordnung von Begriffen, von Wichtigerem und Unwichtigerem, von Abstraktem und Konkretem zu. Die höhere Abstraktionskompetenz ermöglicht die Sondierung von Subjektivem und Allgemeinem und den funktionalen Einsatz von Sprache in längeren Textteilen. Die Textkompetenz als ein komplexes Bündel von kognitiven, von Sprachbasis- und Sprachhandlungsfähigkeiten wird zu einem zentralen didaktischen Ziel des Muttersprachunterrichts. Gleichzeitig wächst die Fähigkeit, Sprache in der Interaktion gezielt einzusetzen, als Mittel der Identifikation, z.B. in der eigenen Gruppe oder zur Abgrenzung von der Welt der Erwachsenen. Die Übergangszeit von der Mittel- zur Oberschule, parallel zur Zeit der sog. zweiten Adoleszenz, spiegelt sich in verändertem Lese- und nonverbalem Verhalten wie im Sprachgebrauch wider. Altersspezifische Varietäten sind identitätsstiftend wichtig, Dialogisieren und Argumentieren treten hervor, das Erzählen stagniert. Die Entwicklung der Fähigkeit hin zu einer multiplen, intersubjektiven Sichtweise, die die Strategien des Dialogisierens und Argumentierens, die Empathiefähigkeit und den Besitz eines erweiterten Weltwissens voraussetzt, wird prozesshaft erworben, die Aneignung, besonders in der Schriftlichkeit, beginnt in der Mittelschulzeit, hält aber über die gesamte Jugendzeit an. Wie im Einzelnen die Sprachaneignung geschieht, ist noch nicht vollkommen geklärt, es gilt aber als gesichert, dass eine reichhaltige Interaktion, mit emotionalen Implikationen einerseits, im Umgang mit Mustersprachhandlungen (z.B. im Leseangebot) andererseits auf den Lernenden stimulierend wirkt. Darüber hinaus wissen wir von der kognitiven Sprachforschung, dass die Fähigkeiten zum Bilden von Begriffen, das Abstra-

hieren, Begründen, Zusammenfassen usw. zwar an Sprachkompetenz gebunden sind, sie an sich aber kognitive Fähigkeiten sind, die hinter den Sprachfähigkeiten liegen (vgl. Schmölzer-Eibinger 2008: 44–59).
Das mehrsprachige kommunikative Angebot auch in der Schule spielt besonders beim Zweitsprachlernen eine essentielle Rolle, um so mehr, wenn im Außerschulischen dies nicht selbstverständlich in dem Maße gegeben ist, wie man es in einer zweisprachigen Region wie Südtirol vermuten würde. Der Aufbau einer soliden Sprachbasis, zu der vor allem auch ein reicher Wortschatz gehört, ist ein weiteres wichtiges Ziel. Darüber hinaus braucht es auch Transfermöglichkeiten von der Mutter- in die Zweitsprache im oben genannten kognitiv-akademischen Bereich, welche vor allem für das Erlernen der Schriftsprachlichkeit in der Zweitsprache grundlegend sind. Die wissenschaftliche Forschung zur (ontogenetischen) Schriftsprachentwicklung der Muttersprache (Erstsprache) im Mittelschulalter konzentriert sich derzeit besonders auf die Erforschung der Aneignung von Schreibstrategien und Textsortenkompetenz (vgl. Augst u.a. 2007). Immer wieder wird betont, dass die Entwicklung von sprachlichen Fertigkeiten nicht kontinuierlich, sondern in Schüben und oft gesplittet erfolgt. Dass die Sprachentwicklung darüber hinaus auch vom sozialen Kontext (siehe z.B. das Phänomen der Jugendsprache bis hin zu den Besonderheiten der Geschlechtszugehörigkeit) mit gesteuert wird, ist augenscheinlich. In der Zweit- und Fremdsprachenforschung erhalten Untersuchungen zu den Lernerstufen, zu den Auswirkungen bilingualer Modelle auch auf das Drittsprachlernen und das Sprachverhalten in mehrsprachigen Situationen derzeit große Aufmerksamkeit (vgl. Ahrenholz u.a. 2008).
Zusammenfassend lässt sich also sagen: Für die individuelle Sprachkompetenzentwicklung der Erst- wie der Zweitsprache sind der Aufbau allgemeiner kognitiver Strukturen, die zum Teil transferiert werden können, der Zugang zu Texten (über das Lesen, inklusive Mediennutzung) und die Sprachanwendung im sozial-kommunikativen Kontext extrem wichtig. Darüber hinaus hängt es ein Stück weit von den Neigungen und dem persönlichen Nützen des Lernangebotes ab, welcher individuelle Stil sich bei der Schülerin/ beim Schüler herausbilden wird. So lassen sich auch bei jungen SprachteilhaberInnen Tendenzen zu eher affektiv-kommunikativen, zu informationsorientierten oder synthetisierenden Schreibhaltungen beobachten.
Von den schriftsprachlichen Strategien des Erzählens, Beschreibens/Berichtens, Instruierens und Argumentierens ist zu Beginn der Mittelschulzeit das Erzählen am weitesten entwickelt. Erzählen kann mündlich oder schriftlich erfolgen, eigene oder fremde Erfahrungen inkludieren, aber ebenso im Zusammenhang mit (erzählender) Literatur stehen. Leseerfahrungen – zunehmend auch im medialen Kontext – wiederzugeben und persönlich zu verarbeiten ist für die Schü-

ler/innen motivierend und anregend. Die Strategien des Erzählens, Beschreibens und Berichtens kommen auch in der Zweitsprache zum Tragen und sind im schriftsprachlichen Kontext weitgehend übertragbar.
Bezeichnend für das Mittelschulalter ist auch die sog. materialsystematische Phase (vgl. Augst/Faigel 1986: 116), in der Schüler/innen das reiche Wissensangebot der Sachfächer sowie die zur Verfügung stehenden häuslichen Medien nutzen und in ihre Texte einzubauen lernen. Auch diese Fähigkeit kann sprachübergreifend geschärft werden. Mit Hilfe von gezielten Such-, Lese- und Umformungsstrategien lernen die SchülerInnen im Laufe der Mittelschuljahre ihre Begriffsverwendungen zu schärfen. Das Definieren und der sichere Umgang mit Begriffen, Themenwörtern und Wortverbindungen sind eine Grundvoraussetzung für den anspruchsvolleren Umgang mit Sachtexten, in der Argumentation, in der gestalteten Rede wie in schriftlichen Texten. Interessant ist, dass die SchülerInnen sehr wohl ein Gespür dafür zeigen, dass und wie sie mit formalen Aspekten im Bereich der Textkompetenz, z.B. Vorgaben für Textsorten, umgehen sollen. Auch das Gespür dafür, sich in adressatenbezogenen Texten auf das Gegenüber einzustellen, ist ohne Zweifel präsent, insgesamt handelt es sich aber um einen Prozess, der erst im jungen Erwachsenenalter als abgeschlossen gelten kann. Auch die eigene Meinung ist in dem genannten Lebensalter schon zu vielem global vorhanden, die Fähigkeit zur Differenzierung ist aber noch beschränkt, da der Zugang zu kontroversen Lebensentwürfen meist fehlt und persönliche Erfahrungen in der Regel noch wenig routiniert aufeinander bezogen werden können. In der Mittelschulzeit dieses zu verlangen führte unweigerlich zur Wiedergabe von Plattitüden und Klischees. In jedem Fall hilfreich ist, wenn Themen auf der Grundlage vorliegender Textquellen schriftlich diskutiert werden können. Die Fähigkeit inhaltlich kontroversielle Positionen darzulegen und dazu persönlich Stellung zu beziehen, nimmt – nach dem Schulwechsel – in den ersten Oberschuljahren sprunghaft zu, vorausgesetzt, dass die Schülerin/der Schüler kontinuierlich und in didaktischen aufbereiteten  Lerneinheiten dazu angehalten wird.

## 3. Überprüfung der Schreibkompetenzen am Ende der Mittelschule

Am Übergang Mittelschule-Oberschule sind Jugendliche beim Schreiben in der Muttersprache in der Regel fähig

- zu persönlicher Verarbeitung von Gelesenem;
- zum Extrapolieren und Gewichten von Informationen aus vorliegenden Texten;
- zum Befolgen von Sprachnormen bei mittlerer sprachlicher Komplexität;
- zum produktiven Umgang mit ausgewählten Textsorten.

In Entwicklung befinden sich auch die Kompetenzen

- mit Sprache und Sprachstrategien in einer persönlichen (intersubjektivem) Form umzugehen;
- die eigenen Texte zu überarbeiten und darüber zu reflektieren;
- erworbene Kompetenzen aus einer Sprache in eine andere zu übertragen (Sprachvergleiche);
- Themen argumentativ zu bearbeiten und dazu eigene Haltungen zu formulieren.

Solche Befähigungen sind als erreichte Kompetenzniveaus/Muttersprache am Ende der Mittelschule auch in den Rahmenrichtlinien – mit einer besonderen Betonung der mündlichen Kompetenzen – festgehalten (vgl. Rahmenrichtlinien 2009). Die Befähigungen zum kompetenten Umgang mit Sprache sind nicht nur reproduzierend oder auf Wissenszuwachs ausgerichtet, sondern sie sind auch prozessualer und methodischer Art bzw. sie tangieren Werthaltungen. Neben der Aneignung von Welt- und Fachwissen ist die Förderung von Haltungen und Fertigkeiten von zentraler Bedeutung. Daraus ergibt sich, dass die Einschätzung des persönlichen Bildungsweges bei Prüfungen ebenso eine Rolle spielt. Standardisierung und Vergleichbarkeit einerseits, langfristig angelegte individuelle Förderung andererseits in Balance zu halten, verlangt den Lehrpersonen viel ab. Nur wenn dies garantiert wird, kann die Fähigkeit zur Selbststeuerung der Lernenden nachhaltig unterstützt werden. Auf Letzteres legen auch die Rahmenrichtlinien Wert:

> „Die Unterrichtstätigkeiten haben die Entwicklung von Lernbewusstheit und Lernkompetenz im Blick, Schülerinnen und Schüler entwickeln die Fähigkeit, die Verantwortung für den Lernprozess selbst zu übernehmen, selbstreflexiv eigene Kommunikations- und Lernstrategien auszubilden und weiter zu entwickeln“ (vgl. Rahmenrichtlinien 2009: 20).

Der Gesetzgeber steckt den Rahmen ab, innerhalb dessen die sprachliche Entwicklung am Abschluss der Mittelschule zu überprüfen und das Leistungspotential dokumentiert werden soll. Es gibt  diesbezüglich eine Reihe von wichtigen Dokumenten.
Solche sind z.B. :

- Richtlinien und Verfahren für die Abschlussprüfung über die Mittelschule (Anlage C zu: Lehrpläne, Stundentafeln und Prüfungsordnung für die Mittelschule in der Provinz Bozen aus dem Jahre 1983);
- Rahmenrichtlinien des Landes für die Unterstufe (2009);

- Rundschreiben vom 14.3.2008, Nr. 32 des italienischen Unterrichtsministeriums (betrifft: Bewertung in der Unterstufe, Abschlussprüfung Mittelschule).

Eine reibungslose Umsetzung der rechtlichen Vorgaben im schulischen Rahmen wird durch den Umstand erschwert, dass noch nicht alle rechtlichen Dokumente aufeinander abgestimmt sind. Zwischen der Entstehung der verschiedenen Dokumente gibt es große zeitliche Abstände und dies führt notgedrungen zu Unterschieden in den Konzepten. Dies gilt im Besonderen für die Prüfungsverordnung (1983!), die mit den jüngeren Rahmenrichtlinien 2009 nicht akkordiert ist. Auch haben die sprachlichen Niveaustufen des Gemeinsamen Europäischen Referenzrahmens für Sprachen (vgl. Europarat 2001) zwar in das Europäische Sprachenportfolio (ESP) (vgl. ESP 2006), welches aber nur empfehlenden Charakter hat, Eingang gefunden, noch nicht aber in die Prüfungsverordnung. Auch die zusätzliche gesamtstaatliche Prüfung aus Deutsch/schriftlich zu den Lernbereichen Textverständnis und Einsicht in Sprache/Grammatik, im Schuljahr 2007/08 zum ersten Mal eingeführt (vgl. Gesetz Nr. 176/2007), ist nicht wirklich mit den übrigen Prüfungsteilen abgestimmt. Ebensolches gilt auch für die Vorgaben zur Formulierung der Aufgabenformate (Themenstellungen) wie für die innere Koordinierung (Ziele, Inhalte) zwischen den Sprachfächern.

### 3.1. Anforderungen der schriftlichen Prüfungen (Mittelschule): Deutsch (Muttersprache), Italienisch (Zweitsprache)

Eine entscheidende Rolle spielt bei den schriftlichen Prüfungen eine sehr gute Themen- bzw. Aufgabenstellung. Bei der Formulierung der Aufgabenstellungen sind im Fach Deutsch bzw. Italienisch/L2 laut Prüfungsvorgaben 1983 folgende allgemeine Angaben zu berücksichtigen:

| **Vorgaben aus der Prüfungsverordnung für die Formulierung der Themen in schriftlichen Prüfungen** | |
|---|---|
| **Fach: Deutsch/Muttersprache** | **Fach: Italienisch/L2** |

| | |
|---|---|
| V1. Darlegung, in der der Schüler seine wirklichen Erfahrungen oder Vorstellungen aus der Phantasie (in Form einer Berichterstattung, eines Tagebuches, eines Briefes, einer Erzählung usw.) ausdrücken kann;<br>V2. Behandlung eines Themas von allgemeinbildendem oder sozialem Charakter, das die Darlegung persönlicher Gedanken ermöglicht;<br>V3. Arbeit über ein Thema aus irgendeinem Fachbereich. | V1. Darlegung von persönlichen Erfahrungen oder von Vorstellungen aus der Phantasie in Form einer Berichterstattung, eines Tagebuches, eines Briefes, einer Erzählung usw.;<br><br>V2. Bericht über ein Thema, das in der Zweitsprache oder auch in Zusammenhang mit anderen Fachbereichen behandelt wurde;<br>V3. Nachgestaltung eines gelesenen Textes mit persönlichen Überlegungen zum Inhalt. |

Tab. 1 entnommen aus: Lehrpläne, Stundentafeln und Prüfungsordnung für die Mittelschule in der Provinz Bozen. (1983)

Die Gegenüberstellung in Tabelle 1 zeigt, dass für die expressiv-erzählende Themenstellung in beiden Sprachen die selben Textsorten in Betracht gezogen werden, z.B. Tagebuch, Brief, Erzählung (siehe D_V1, I_V1). In den Themenstellungen V2 und V3 werden nicht Textsorten, sondern nur Schreibhaltungen genannt: „darlegen" (D_V2), „berichten" (I_V2) oder die Textsorte bleibt offen, z.B. „Arbeit über…" (D_V3). In beiden Fächern ist ein fachübergreifendes Thema verpflichtend vorgesehen (siehe D_V3 und I_V2). Der wesentliche qualitative Unterschied die Textsorte und die Sprachstrategien betreffend besteht zwischen der Themenstellung D_V2 und I_V3: In der deutschen Prüfung wird ein Thema allgemeinbildender oder sozialer Art vorgegeben, zu dem die Schülerin/der Schüler „persönliche Gedanken" darlegen soll, was einer Vorstufe des Argumentierens gleichkommt. „Persönliche Überlegungen" werden auch in Italienisch erwartet, allerdings in engem Zusammenhang mit einem gelesenen Text. In I_V3 steht die Nachgestaltung eines gelesenen Textes im Mittelpunkt, was bedeutet, dass mit der schriftlichen Arbeit auch die Lesekompetenz kontrolliert wird. Letzteres wird in der Muttersprache seit dem Schuljahr 2007/08 in der zusätzlichen gesamtstaatlichen Prüfung eingefordert (vgl. Gesetz vom 25.10.2007, Nr. 176). Damit wird umgesetzt, was in neueren schreibdidaktischen Konzepten auch für die Muttersprache angemahnt wird, nämlich, die Bedeutung des Lesens für den Schreibakt selbst aufzuwerten. Um den Lese- und Schreibprozess zu portionieren und die Leistungen der Schreibenden besser ergründen zu können, wird zunehmend auch in muttersprachlichen Prü-

fungsvorgaben höherer Schulstufen von Textvorlagen, die lesend bearbeitet werden müssen, ausgegangen.[3]

Der Gesetzgeber gibt zwar Textsorten/Schreibhaltungen für die schriftlichen Prüfungen vor, belässt es aber den Prüfungskommissionen[4] die Aufgabenstellungen zusammen zu stellen. Es gibt für die FachlehrerInnen bzw. die Kommission keine Hinweise zur formalen Konstruktion der Themenstellungen bzw. zur Formulierung eines erwünschten Erwartungshorizontes. Der knappe Verweis auf Kriterien für die Beurteilung[5] verrät nichts über Gewichtungen derselben oder über einen eventuellen Bezug zu Niveaustufen in der Bewertung. Eine innere Koordinierung der Ansprüche an die muttersprachliche und zweitsprachliche SchülerInnenleistung wird zwar in Hinweisen zu gemeinsamen Aspekten (z.B. die Betonung von Textsorten/Schreibhaltungen; Verweis auf lebensweltliche Anbindung) oder aber durch koordinierte Abstufungen (z.B. Tiefe der Verarbeitung eines Sachthemas: „Arbeit über…“(D3) versus „Bericht über…“(I 2)) sichtbar. Ein Korrekturschlüssel oder feinjustierte Kriterienkataloge, die z.B. auf bestimmte Textsorten Rücksicht nehmen und ebenso Ausdruck einer gemeinsamen Prüfungskultur sind, fehlen hingegen. Die Korrektur und die Bewertung nehmen die FachlehrerInnen in Unterkommissionen bzw. nimmt die gesamte Prüfungskommission vor. Sie beziehen sich dabei auf einen intern beschlossenen Kriterienraster, der sich mit den Kriterien in der curricularen Planung der Schule deckt. An eine externe Zweitkorrektur o.Ä. ist nicht gedacht.

## 3.2. Beispiele zu Themenstellungen aus der Prüfungspraxis

Einige Beispiele zu Prüfungsthemen aus dem Schuljahr 2006/07 einer zufällig ausgewählten Klasse sollen zeigen, wie die (unterschiedlichen) Anforderungen

3 Vergleiche z.B. die verpflichtenden Textvorlagen in Themenstellung A, B bei der zentralen italienischen staatlichen Abschlussprüfung der Sekundarstufe II, Fach Muttersprache (vgl. Ministerialdekret vom 23.04.2003, Nr. 41) wie auch in der beschlossenen Reform der österreichischen Reifeprüfung aus Deutsch/AHS. www.bifie.at (Stand: 25.10.10)

4 Die Prüfungspräsidentin/der Prüfungspräsident wird von außen berufen, die weiteren Mitglieder der Kommission sind Lehrer/innen der Klasse.

5 Siehe dazu die einschlägigen gesetzlichen Textstellen zu Deutsch/Muttersprache: „Die schriftliche Arbeit hat zu überprüfen, ob das Thema inneren Zusammenhang und logische Gedankenfolge aufweist und ob der Schüler über persönlichen Ausdruck und normgerechten und angemessenen Sprachgebrauch verfügt.“ Bezüglich Italienisch/L2 heißt es: „Die schriftliche Arbeit hat den normgerechten und angemessenen Sprachgebrauch des Schülers zu überprüfen, ebenso die Fähigkeit, ob er eigene Gedanken klar ausdrücken kann.“ (vgl. Lehrplan (1983), Auszug aus den Prüfungskriterien)

an die schriftlichen Mittelschulprüfungen in Deutsch und in Italienisch/L2 umgesetzt werden.[6]

| Beispiel: Themenstellungen einer Prüfungskommission | | |
|---|---|---|
| **Deutsch/Muttersprache** | **Italienisch/L2** | **Englisch/L3** |
| In Gedanken achtete ich viel zu wenig auf das, was um mich herum passierte… (11x)[8] | Sei in soffitta e trovi un vecchio baule. Senti dei rumori all'interno. Con un po' di paura lo apri e…Racconta! (4x) | (mit Textvorlage)<br>Füge das korrekte Subjekt, Objekt und possessiv Pronomen ein.<br>Bilde die Mehrzahl der Namen, die in der Klammer sind. |
| Was sind aus deiner Sicht sinnvolle Formen der Freizeitbeschäftigung? Führe einige an und erläutere sie möglichst anschaulich! (4x) | Scrivi una lettera a un'amica e racconta di una gita scolastica che ti è piaciuta particolarmente. Descrivi i luoghi che hai visitato, le persone che hai incontrato e scrivi che cosa ti è piaciuto e perché! (4x) | Setze die Verben im *present simple* ein.<br>Setze den richtigen Komparativ und Superlativ ein.<br>Bilde Sätze in der Negativform und Fragesätze.<br>Setze das richtige Fragewort ein. |
| Die Industrielle Revolution hat unsere Welt entscheidend verändert. Lege dein Wissen darüber möglichst anschaulich dar! (2x) | Scrivi di un personaggio in italiano, che ti ha particolarmente colpito! (9x) | Wähle die richtige Präposition, um den Satz zu vervollständigen.<br>Eine Geistergeschichte.<br>Setze die Verben in Klammer im *past simple* ein und vervollständige den Text.<br>Bilde korrekte Sätze (Satzelemente sind vorgegeben).<br>Setze die korrekte Zeit ein: *present simple, present progressive, past simple, going to future and will future.* Achte auf bestimmte Wörter, die dir die Wahl erleichtern.<br>Bilde korrekte Sätze mit den Adverbien in den Bildern.<br>Schreibe eine kurze Geschichte mit den Wörtern im Box (10 Sätze). |

Tab. 2

Die Aufgabenformate in den beiden Sprachfächern Deutsch und Italienisch sind in sich unterschiedlich: einige sind zweigeteilt und enthalten im zweiten Teil explizite oder implizite Hinweise zur Textsorte oder zu den Schreibhal-

6 Ergänzend wird in Tabelle 2 auch die dreigeteilte Aufgabenstellung aus Englisch/L3 derselben Klasse angeführt. In dem vorliegenden Fall wurde in Englisch von allen SchülerInnen dieselbe Aufgabenstellung ausgeführt. Die Prüfungsverordnung sieht vor, dass jede Aufgabenstellung in Englisch/L3 sich in verschiedene Teilaufgaben gliedert, die auf die Kompetenzbereiche: a) Textverständnis, b) Wortschatz, c) Länderkundliches Wissen eingehen müssen (siehe dazu Prüfungskriterien, Lehrplan (1983)).

8 Die Zahlen in Klammern drücken aus, wie oft ein Thema von den SchülerInnen der betroffenen Klasse gewählt wurde.

tungen (z.B. D 2, I 2). Andere Themen sind so formuliert, das der Schülerin/dem Schüler Schritte zur inhaltlichen Abfolge nahe gelegt werden (siehe I 1). Besonders das Thema D_1 ist aber so offen formuliert, dass für die SchülerInnen Missverständnisse provoziert werden.
Es ist für den Leser nicht von vorne herein klar, ob es sich bei der Themenstellung D 1 um einen Textanfang oder um einen Kern- oder Lehrsatz einer zu gestaltenden Geschichte handelt. Da jede pädagogisch/didaktische Hilfestellung zu Textsorte/Schreibhaltung fehlt, führte die Angabe in den Schülertexten zu unterschiedlichsten Umsetzungen, von denen kaum eine von der korrigierenden Lehrperson goutiert wurde. Hinter der Themenstellung stand – den Korrekturen nach zu schließen – die Erwartungshaltung, dass der Satz der Themenstellung inhaltlich und formal Auslöser des Höhepunktes der Erzählung sein, der entsprechend ausgestaltet werden sollte. Dass der „gedankenverlorene" Ich-Erzähler gleichzeitig Hauptakteur/in sein sollte, erforderte einen Spannungsbogen, den die meisten Schüler/innen nicht schafften.[9]

Die Schüler/innen wählten die zur Wahl stehenden Themenstellungen aus Deutsch unterschiedlich oft: Am häufigsten wurde auf die erste Themenstellung zugegriffen, wohl auch deshalb, weil die beiden sich implizit anbietenden Textsorten: Erzählung, Phantasiegeschichte seit langem und schulstufenübergreifend eingeübt werden konnten. Auffallend selten wurde das Thema D 3 gewählt, auch wenn der Inhalt fester Bestandteil im Geschichteunterricht einer 8. Schulstufe ist. Der inhaltliche Aspekt der Themenstellung im ersten Teil ist weit und undifferenziert, wogegen die Anleitung im zweiten Teil („Lege dein Wissen drüber möglichst anschaulich dar.") der Schreibenden/dem Schreibenden a) viel (Spezial) Wissen und b) die Darlegung in einer möglichst „anschaulichen" Sprache und Textform abverlangt. Diese Kombination an Ansprüchen – hier Informationsdichte, dort „Anschaulichkeit" – stellt als Basiskompetenz, die von allen SchülerInnen bewältigt werden soll, eine Überforderung dar.

9 Einige Beispiele:
S1 endet mit einem persönlichen Vorsatz: „Dieser Vorfall war mir eine Lehre und ich werde nun besser aufpassen, was um mich herum passiert."
S 2 erzählt eine Geschichte, in dem ihm/ihr vor lauter Unachtsamkeit beinahe ein schwerer Skiunfall passiert wäre. Er/sie verpackt die inhaltliche Aussage der Themenstellung in die erfolglosen Anweisungen des Vaters: „Immer wieder schrie der Vater mir zu, dass ich mehr aufpassen sollte aber ich tat es nicht."
S3 berichtet von einem Tag, an dem ihm/ihr zumindest dreimal – weil gedankenverloren- Ungeschicklichkeiten passierten, was zu zumindest drei „Höhepunkten" in der Geschichte führte. Der Text endet schließlich mit dem Satz: „Doch dieses ganze Theater alles nur, weil ich mit dem Kopf in den Wolken bin, und so ein Tag ist darum bei mir keine Seltenheit."
S 6 kommentiert mitten im Text das Verhalten der Hauptfigur Alex: „Doch Alex konnte heute einfach nicht mitlachen, denn er war mit den Gedanken nicht bei Peter und Franz, sondern bei Jasmin.", was die korrigierende Lehrperson kommentierte mit: „Der Höhepunkt kommt nicht gebührend zur Geltung".

In der Zweitsprache hingegen ist das fachbezogene Thema das am häufigsten gewählte (9/17), da es sich auf ein begrenztes schulisches Wissen bezieht, welches im Fach selbst behandelt worden ist. Dass das Thema eigentlich nicht der gesetzlichen Vorgabe (siehe I_V 3) entspricht, welche die „Nachgestaltung“ eines Textes nahelegt, muss hier für die Schüler/innen kein Nachteil gewesen sein.
Die Themenstellungen sind unterschiedlich formuliert und was die Erwartungshaltung anbelangt nicht immer eindeutig genug – weder für die Schüler/innen noch für die mitkorrigierenden Lehrpersonen. Aufgabenformate, in denen die Wiedergabe und Verarbeitung von geordnetem Sachwissen im Vordergrund steht und mit denen erst in den Mittelschuljahren umgegangen wird, bedürfen vielfältiger Übungssituationen im vorangegangenen Unterricht. Nur nach wiederholtem Üben in stressfreien Situationen wagen sich die Schüler/innen in Prüfungssituationen in der Regel an solche Themen heran.

### 3.3. Verbale Beurteilungen von Schülerleistungen: Deutsch (Muttersprache), Italienisch/L2 – Beispiele

Die Prüfungsarbeiten werden von einer Unterkommission korrigiert, was in der Praxis aber nicht immer bedeutet, dass die Lehrer/innen dabei gleichzeitig über ein und demselben Text sitzen. Die Bewertung des den Autorinnen vorliegenden Prüfungsmaterials, das aus dem Schuljahr 2006/07 stammt, erfolgte mittels einer verbalen Beurteilung, die dann von der Prüfungskommission beschlossen wurde.
Verbale Beurteilungen sind sprachlich ausgeführte Bewertungen, die eine Ziffernnote oder eine Punktezahl ersetzen. Verbale Beurteilungen können aber auch als Ergänzung zu den zu vergebenden Punkten oder Ziffernnoten auftreten, wie in der Abschlussprüfung der Oberschule schon seit der gesamtstaatlichen italienischen Reform 1998/99 oder in der Mittelschule seit der Reform 2007/08 gegeben (vgl. Rundschreiben 14.3.2008, Nr. 32).
Die verbale Bewertung erfüllt wie die Ziffernote eine „bewertend-prüfende“ Funktion. Sie kann aber so angelegt werden, dass sie zusätzlich „fördernd“ wirkt und der Schreibenden/dem Schreibenden signalisiert, welche Aspekte ihrer/seiner Schreibkompetenz wie weiterentwickelt werden sollen. Die „fördernde“ Funktion greift individuell angelegte Aspekte der Schreibentwicklung auf und vermeidet eine zu starke Defizitorientierung. Die Lehrperson empfiehlt sich gleichsam als Berater/in für Schreibentwicklung und verweist auf den erfolgreichen Einsatz von Strategien, eventuell auch aus der Optik einer gemeinsamen Sprachenerziehung. Diese fördernde Haltung muss bei der Ab-

schlussprüfung nicht ausgeschaltet werden, zumal ja Ziel der Prüfung selbst auch ist, die erreichten Kompetenzen dezidiert zu benennen (vgl. Landesgesetz v. 16.7.2008, Nr.5, Anlage, Ar.16, Abs.5). Solche Kompetenzbeschreibungen können der Kandidatin/dem Kandidaten bei der Entscheidung über den weiteren Bildungsweg helfen, jedenfalls wird in den Rahmenrichtlinien *Kompetenz* in diesem Sinne verstanden:

„Laut Empfehlung des Europäischen Parlaments und des Europäischen Rates vom 7. September 2006 wird Kompetenz definiert als nachgewiesene Fähigkeit, Kenntnisse, Fertigkeiten sowie persönliche, soziale und/oder methodische Fähigkeiten in Arbeits- und Lernsituationen und für die berufliche und/oder persönliche Entwicklung zu nutzen." (Rahmenrichtlinien 2009: 18)

Die folgenden Beispiele zeigen uns, welche verbalen Beurteilungen die SchülerInnen S 1, S 2, S 9 bzw. S 10 erhalten haben.

| **Verbale Beurteilungen aus Prüfungsbeispielen** | |
|---|---|
| **Deutsch/Muttersprache** | **Italienisch/L2** |
| Bsp.: (S1/Thema: 1)<br>Sprachlich ist die Arbeit bis auf zwei Stellen klar und angemessen. Der Aufbau ist gut und steigert sich einem Höhepunkt entgegen, der dann nicht ganz dem Thema entspricht. | Es.: (S1/tema nr. 1)<br>Ha rispettato la traccia. Il contenuto è sufficiente, la forma abbastanza scorrevole. Il lessico è approssimativo. Errori nell'uso dei tempi verbali e nell'uso delle preposizioni. |
| Bsp.: (S9/Thema: 2)<br>Thema erfasst und gründlich ausgearbeitet, klare, nachvollziehbare Gedankengänge, sicherer Ausdruck, sichere Beherrschung der Sprachnormen, einige Generalisierungen. | Es.: (S2/tema nr. 2)<br>Ha rispettato la traccia. Il contenuto è sufficiente, la forma abbastanza scorrevole. L'esposizione è comprensibile e il lessico abbastanza appropriato. Errori nell'uso dei tempi verbali, preposizioni e nomi plurali. |
| Bsp.: (S10/Thema: 3)<br>Das Thema ist sehr weitläufig und erschöpfend behandelt. Auch Persönliches findet sich im Inhalt. Sprachlich sind einige Stellen etwas ungeschickt ausgeführt. Der Aufbau klar und folgerichtig. | Es.: (S9/tema nr. 3)<br>Ha rispettato la traccia. Il contenuto è buono, la forma scorrevole. L'esposizione è chiara, il lessico ricco e appropriato. Molte le informazioni su Schindler e Anna Frank. Qualche errori di grammatica: uso tempi verbali e preposizioni. |

Tab. 3 (Legende: Schüler/in = S; die Nummer des Themas bezieht sich auf Themenstellung D 1, D 2, D 3 oder I 1, I 2, I 3, siehe Tabelle 2)

Die unpersönliche Art der verbalen Bewertungen (Deutsch wie Italienisch)[10] unterstreicht die betont legitimierende und Output orientierte Absicht der Formulierungen. Die Formulierungen bilden eine Zäsur am Ende der Schulstufe ab. Dabei ist die Sprache formelhaft, es werden immer wieder dieselben bewertenden Ausdrücke verwendet: z.B. „klar“, „angemessen“, „sicher“; „chiaro“, „scorrevole“, „appropriato“. Diese deuten auf eine implizit vorhandene Abstufung von Niveaustufen hin (z.B. der „Aufbau“ ist. „klar“/“logisch“ versus „verworren“/“wenig durchdacht“). Sicher zu kurz bzw. zu undifferenziert sind wertende Ausdrücke dieser Art aber, wenn sie über alle Textfunktionen summativ darüber gelegt werden, z.B. auch über Teile, in denen persönliche Gestaltung und Meinung gefordert sind. Der Aufbau der verbalen Beurteilung folgt in Deutsch meist der „klassischen Dreiteilung“: Inhalt/Bezug auf Themenstellung – Aufbau – Sprache, in Italienisch einer etwas differenzierteren, aber stark sprachnormorientierten Unterteilung: traccia, contenuto, forma, esposizione, lessico, grammatica. Die Gewichtung der verschiedenen impliziten Kriterien ist nicht nachvollziehbar, es sei denn die Unterteilung von Sprachnormverstößen wird als Indiz für „besonders wichtig“ genommen. Auffällig ist auch, dass die Beschreibung der inhaltlichen Gestaltung und Vertiefung sehr unterschiedlich angegangen wird. Meist bleibt es bei der Einschätzung, ob die vorgegebene Themenstellung erfasst wurde: Beispiele: „Ha rispettato la traccia“ oder „Themenstellung nicht berücksichtigt“.
Bei den verbalen Beurteilungen sind verschiedene Korrektoren federführend, die, obwohl in einer Unterkommission zur Zusammenarbeit verpflichtet, nicht immer mit denselben Differenzierungen vorgehen.
So wie schon in den Themenstellungen sind auch in den Lehrerkommentaren eindeutige Aussagen zur Umsetzung der Textsorte, zu Adressatengemäßheit, persönlichem Stil selten, dies gilt z.B. für die Textsorte Brief in I_2 wie für die Erzählung in I_1. Demgegenüber wird das Textsortenmerkmal „Hinführen zum Höhepunkt“ in einer Erzählung – wie bereits weiter oben erwähnt - in den Schülertexten zu D 1 vehement eingefordert. Bei 10 von 11 Lehrerbeurteilungen zum Thema D 1 wird vermerkt, dass der Höhepunkt der Erzählung nicht angemessen getroffen worden sei.
Die korrigierenden Lehrer/innen legen den Arbeiten zwar Kriterien zugrunde, da diese in sich aber nicht differenziert sind, kommt es in den verbalen Begründungen zu schematisierten Einschätzungen; differenzierte Vermerke werden manchmal, aber nicht durchgehend gemacht. Beobachtungen zur Sprachbasis überwiegen gegenüber Beobachtungen zu kognitiven Leistungen. Die

10 In dem vorliegenden Fall standen uns die verbalen Beurteilungen der FachlehrerInnen, aber leider nicht die Kriterien zur Verfügung, wenngleich sie sicher erstellt worden sind. Dieser Umstand erschwert die Einschätzung der verbalen Beurteilungen (Lehrerkommentare) durch die Autorinnen.

Würdigung der kognitiven Leistungen ist für die Schreibentwicklung über die Schulstufe hinaus aber sehr wichtig.

## 4. Prüfungskultur und Schnittstellendidaktik

Abschlussprüfungen sind ein Anlass, an dem nicht nur Einzelleistungen gemessen werden, sondern die Ganzheitlichkeit der erworbenen Bildung auf dem Prüfstand steht. Der Grundstein für die komplexe Vernetzung der Fähigkeiten im Kopf der Schülerin/des Schülers wird in einer Lehr- und Lernhaltung angelegt, die der Anbindung von Wissen und Erfahrungen, dem Transfer von Fähigkeiten und der Stärkung kognitiver Kompetenzen Raum und Zeit gibt. Transparenz und innere Koordinierung der Bildungsziele, -inhalte und -anforderungen sind konstitutive Voraussetzungen bei Prüfungen, unabhängig davon, ob diese mehr „individualorientiert" oder „feststellungsorientiert" sind, ja, sie sind notwendig, damit Fördern und Prüfen nicht als Gegensatz erlebt werden. Nahtstellen wie der Übergang von einer Schulstelle zur nächsten haben etwas Abschließendes wie etwas Vorausblickendes an sich. In diesem Spannungsfeld kann sich Nachhaltigkeit entwickeln.
Die LehrerInnen der Schule, die uns ihre Unterlagen zur Verfügung gestellt haben, zeigen großes Bemühen, den Prüfungsvorgaben gerecht zu werden. Durch eine ausgewogene Differenzierung in den Kriterien (Niveaustufendifferenzierung), einer stärkeren Beachtung der kognitiven Kompetenzen und der Abstimmung derselben auch zwischen den Sprachfächern könnte ein nächster Qualitätsschritt erreicht werden. Der Weg dahin, dass die Schülertexte aus Deutsch und Italienisch derselben Schüler/innen als eine kognitive Leistung in zwei Sprachen gesehen werden, wäre dann nicht mehr weit.
Eine Reihe von Südtiroler Schulen haben in den letzten beiden Jahren ihre Kriterien weiterentwickelt. Es gibt vielfältige Beispiele, in denen das Bemühen um eine individualisierende Kompetenzenorientierung und um eine disziplinübergreifende Zielgebung in den Sprachfächern inzwischen gut sichtbar ist.[11]

Einer neuen Prüfungskultur geht die Weiterentwicklung des didaktischen Verhaltens im alltäglichen Unterricht voraus. Dieses baut auf an Ganzheitlichkeit und Interdisziplinarität ausgerichteten Grundsätzen von Sprachbildung auf:

- Der Umgang mit Niveaustufen bei den Kriterien auch zu muttersprachlichen Leistungen erleichtert den professionellen Austausch unter den Muttersprach-, Zweit- und Fremdsprachenlehrer/innen.

11 Als Beispiele seien die Kriterien der Schulsprengel Naturns und Brixen-Milland genannt.

- Die Stärkung einer lernerorientierten beratungs- und kriterienorientierten Prüfungskultur fördert gleichzeitig die Befähigung zu fachlicher Selbsteinschätzung und zu einem autonomen wie kooperativen Arbeiten der Schülerin/des Schülers.
- Verbale Schülerbeurteilungen, die auf eine Dialogsituation hin formuliert werden und den Aspekt der Rückmeldung antizipieren, tragen zur Nachhaltigkeit pädagogischen Handelns bei.

Abschlussprüfungen geben mehr von sich preis, als vordergründig angenommen: Abschlüsse dokumentieren nicht nur die erreichten Leistungen der Schüler/innen wieder, sie geben vielmehr auch Auskunft, wie die Schule sich im Bildungsweg der Jugendlichen positioniert und wie der Übergang auf die nächste Schulstufe vorbereitet wird. Es ist den Herkunftsschulen zu wünschen, dass der Einsatz ihrer Lehrpersonen in der Nachfolgeschule erkannt wird und nicht jede Bildungsstufe von sich behauptet, das „Rad neu erfinden zu müssen".

## Literaturangaben

Ahrenholz, Bernt / Oomen-Welke, Ingelore (Hg.) (2008): Deutschunterricht in Theorie und Praxis (DTP), Bd. 9: Deutsch als Zweitsprache. Baltmannsweiler: Schneider Verlag Hohengehren

Augst, Gerhard u. a. (2007): Text – Sorten - Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter. Frankfurt: Peter Lang

Augst, Gerhard / Faigel, Peter (1986): Von der Reihung zur Gestaltung. Untersuchungen zur Ontogenese der schriftsprachlichen Fähigkeiten von 13 - 23 Jahren. Frankfurt: Peter Lang.

Becker-Mrotzek, Michael / Böttcher, Ingrid (2006): Schreibkompetenz entwickeln und beurteilen. Praxishandbuch für die Sekundarstufe I und II. Berlin: Cornelsen Scriptor

Europarat (2001): Gemeinsamer Europäischer Referenzrahmen für Sprachen. München: Langenscheidt

Europäisches Sprachenportfolio (2006): Für Schülerinnen und Schüler von 11 bis 14 Jahren. Bozen: Autonome Provinz Bozen-Südtirol

Gelmi, Rita / Fischer, Maria Luise / Saxalber, Annemarie (1992): Lernzielvergleich Muttersprache - Zweitsprache unter dem Blickpunkt gemeinsamer Spracherziehung. In: Gelmi, Rita / Saxalber, Annemarie (Hg.): Integrierte Sprachdidaktik: Muttersprache – Zweitsprache. Theoretische Beiträge. Bozen. 22-44 (= Beiträge zu Er-

ziehung und Unterricht in Südtirol. Schriftenreihe des Pädagogischen Institutes Bozen)

Grießhaber, Wilhelm: Sprachstanddiagnose im kindlichen Zweitspracherwerb: Funktional - pragmatische Fundierung der Profilanalyse. In: http://spzwww.uni-muenster.de/~griesha/pub/tprofilanalyse-azm-05.pdf (Stand: 30.1.10)

KOLIPSI: Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung. www.eurac.edu/de/research/projekts/Projekt/Details.aspx?pid=1818 (Stand: 29.5.10)

Schmölzer-Eibinger, Sabine (2008): Lernen in der Zweitsprache. Grundlagen und Verfahren der Förderung der Textkompetenzen in mehrsprachigen Klassen. Tübingen: Gunter Narr Verlag

Ruf, Urs / Winter, Felix (2006): Qualitäten finden. In: Diagnostizieren und Fördern. Stärken entdecken – Können entwickeln. Friedrich Jahresheft XXIV. Seelze: Friedrich, 56-59

## Gesetzliche Unterlagen

Bestimmungen über Schlussbewertungen und Prüfungen an Grund-, Mittel- und Oberschulen staatlicher und nichtstaatlicher Art der Autonomen Provinz Bozen – Südtirol (Ministerialverordnung v. 21.5.2001, Nr. 90)

Landesgesetz v. 16.7.2008, Nr. 5. Festlegung allgemeiner Kriterien im Bereich der Bewertung der Schülerinnen und Schüler an Grund- und Mittelschulen des Landes.

Bestimmungen im Bereich der Bewertung der Schülerinnen und Schüler an Oberschulen des Landes. (Beschluss der Landesregierung Nr. 2485 v. 12.10.2009)

Legge di conversione vom 25.10.2007, Nr. 176 (Betrifft: gesamtstaatliche Prüfung)

Lehrpläne, Stundentafeln und Prüfungsordnung für die Mittelschule in der Provinz Bozen. Landesgesetz vom 6. Dezember 1983, Nr. 48. Anlage C: Richtlinien und Verfahren für die Abschlussprüfung über die Mittelschule in der Schule mit deutscher und italienischer Unterrichtssprache der Autonomen Provinz Bozen

Ministerialdekret v. 23. April 2003, Nr. 41(Betrifft: Verordnung über die Durchführung der ersten und zweiten schriftlichen Prüfung bei der Abschlussprüfung der Oberschulen für das Schuljahr 2002/2003)

Neue (österreichische) schriftliche Reifeprüfung Deutsch/AHS. http://www.bifie.at (Stand: 25.10.10)

Rahmenrichtlinien des Landes für die Festlegung der Curricula für die Grundschule und die Mittelschule an den autonomen deutschsprachigen Schulen in Südtirol (Beschluss der Landesregierung v. 19.1. 2009, Nr. 81)

Rundschreiben 14.3.2008, Nr. 32 des italienischen Unterrichtsministeriums (Betrifft: Bewertung in der Unterstufe, Abschlussprüfung Mittelschule)

# Autorenbiographien

**Andrea Abel**
(<andrea.abel@eurac.edu>) ist Sprachwissenschaftlerin und seit 1997 wissenschaftliche Mitarbeiterin, seit 2004 Koordinatorin des Instituts für Fachkommunikation und Mehrsprachigkeit der Europäischen Akademie Bozen. Zudem ist sie Lehrbeauftragte an den Universitäten Bozen und Trient. Ihre Forschungsschwerpunkte liegen in den Bereichen Lexikographie (u.a. pädagogische Lexikographie, Computerlexikographie), Sprachkontaktforschung (u.a. Sprachvarietäten, Sprachattitüden), Deutsch als Erst- und Zweitsprache (u.a. Erhebung und Beschreibung von Sprachkompetenzen, Gemeinsamer europäischer Referenzrahmen für Sprachen, Migration und Sprachen) und der Korpuslinguistik (u.a. Varietätenkorpora, Lernerkorpora).

**Stefanie Anstein**
(<stefanie.anstein@eurac.edu>) ist Computerlinguistin am Institut für Fachkommunikation und Mehrsprachigkeit der Europäischen Akademie Bozen. Dort ist sie im Bereich Sprachtechnologien tätig und arbeitet seit 2006 hauptsächlich an der Initiative *Korpus Südtirol*. Ihre Forschungsinteressen sind in den Bereichen Korpuslinguistik und Varietätenlinguistik angesiedelt.

**Margit Oberhammer**
(<margit.oberhammer@unibz.it>) absolvierte ein Studium der Germanistik und Kunstgeschichte und ist Mitarbeiterin der Freien Universität Bozen sowie Kulturpublizistin und Kritikerin für verschiedene Medien; Veröffentlichungen in Zeitungen und Zeitschriften.

**Stefanos Petrakis**
(<petrakis@ifi.uzh.ch>) ist Computerlinguist am Institut für Computerlinguistik der Universität Zürich. Er war vorher am Institut für Fachkommunikation und Mehrsprachigkeit der Europäischen Akademie Bozen, Bereich Sprachtechnologien, im Rahmen der Initiative Korpus Südtirol tätig. Seine Forschungsinteressen liegen hauptsächlich in der Korpuslinguistik und der Semantik.

## Cyril Belica

(<belica@ids-mannheim.de>) ist am Institut für Deutsche Sprache in Mannheim für den *Programmbereich Korpuslinguistik* verantwortlich. Seine Forschungsschwerpunkte sind: Korpuslinguistik, Methodik der prämissenarmen Analyse und linguistischen Interpretation von sprachbezogenen Erwartbarkeiten, auf Theoriebildung zielende Generalisierungsstrategien im Lexikon-Syntax-Kontinuum, wissenschaftsmethodische Grundlagen empirischer Sprachforschung, lexikalische Semantik.

## Renata Zanin

(<renatazanin@unibz.it>) ist seit 2005 Forscherin und Dozentin für Deutsche Sprache und Übersetzung an der Freien Universität Bozen. Ihre Forschungsinteressen liegen im Bereich der Übersetzung und der Sprachdidaktik Deutsch als Zweit- und Fremdsprache.

## Hans Drumbl

(<jdrumbl@unibz.it>), Professor für Deutsche Sprache an der Freien Universität Bozen (Fakultät für Bildungswissenschaften, Brixen) hat seit 1969 an den Universitäten Bergamo, Florenz, Triest, Modena, und Reggio Emilia sowie an der Katholischen Universität Mailand in den Bereichen Deutsche Sprache, Literatur, Philologie unterrichtet. Arbeitsgebiete: literarische Hermeneutik, Mehrsprachigkeit und Übersetzen im Mittelalter und in der Gegenwart, Spracherwerb und Sprachdidaktik im Umfeld von Deutsch als Fremd- und Zweitsprache.

## Annemarie Saxalber

(<annemarie.saxalber@unibz.it>), Prof.in für Deutsche Sprache und ihre Didaktik an der Bildungswissenschaftlichen Fakultät der Freien Universität Bozen. Von 2007–2010 Professorin am Österreichischen Kompetenzzentrum für Deutschdidaktik an der Universität Klagenfurt. Forschungsschwerpunkte: Schreibforschung und -didaktik, Sprachforschung und -förderung in einem mehrsprachigen Gebiet, Integrierte Sprachdidaktik, Leistungsüberprüfung im Sprachunterricht.

### Alexander Geyken

(<geyken@bbaw.de >), arbeitet seit 1999 an der Berlin-Brandenburgischen Akademie der Wissenschaften im Projekt *Digitales Wörterbuch*, ist dort Teil der Steuerungsgruppe und hat maßgeblich am Aufbau der Webseite www.dwds.de mitgewirkt. Seine Forschungsschwerpunkte sind Computerlexikographie, Morphologie, Korpuslinguistik und allgemein die Wortschatzentwicklung der deutschen Sprache.

### Rita Gelmi

(<rgelmi@bz.it>), Studium der Fremden Sprachen und Literaturen (Deutsch, Englisch) in Urbino, Italien, lebt und arbeitet in Bozen. War Inspektorin für Italienisch Zweite Sprache an den deutschsprachigen Schulen in Südtirol. Koordinierte das Projekt *Europäisches Sprachenportfolio* für die Schulen in Südtirol und hat die Thematik *Gemeinsame Sprachendidaktik* besonders vertieft. Lehraufträge an der Freien Universität Bozen.

Die Vermittlung von Wissen steht heute vor einer zweifachen Herausforderung: dem Anspruch auf Verständlichkeit trotz hoher Komplexität der Inhalte und dem Erreichen akzeptabler Ergebnisse in der Schule in Situationen von immer größerer Diversität. Daraus ergeben sich für die Sprachwissenschaft neue Schwerpunkte: eine konstante Besinnung auf die Phänomene, das sind die kommunikativen Akte von Mitgliedern einer Sprachgemeinschaft im Alltag und in Bereichen fachlicher Spezialisierung, das sind aber auch variable Kontexte der Sprechakte, die eine Koorientierung der sozialen Akteure verlangen, die ebenso gelernt und geübt werden muss, wie die Sprachen in ihren formalen und strukturellen Eigenheiten.
In diesem Band werden Forschungen und theoretische Ansätze vorgelegt, die ausgehend von diesen Prämissen Korpusrecherchen für die Erforschung und für die didaktische Vermittlung der drei Schulsprachen Deutsch, Italienisch und Englisch nutzbar machen. Besondere Schwerpunkte sind Varietäten der deutschen Sprache in Südtirol und die Interface-Problematik.







30,00 Euro

www.unibz.it/universitypress